

**ADVANCED PROTEOMIC CHARACTERIZATION OF THE 26S PROTEASOME IN
ARABIDOPSIS REVEALS INSIGHTS INTO COMPOSITION AND ASSEMBLY**

by

David C. Gemperline

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: TBD

The dissertation is approved by the following members of the Final Oral Committee:

Richard D. Vierstra, Professor, Genetics

Richard Amasino, Professor, Biochemistry

Josh Coon, Professor, Chemistry

Donna Fernandez, Professor, Botany

Patrick Masson, Professor, Genetics

© Copyright by David C. Gemperline 2016

All Rights Reserved

To Erin, my wife.

ACKNOWLEDGMENTS

Science doesn't purvey absolute truth. Science is a mechanism, a way of trying to improve your knowledge of nature. It's a system for testing your thoughts against the universe and seeing whether they match. This works not just for the ordinary aspects of science, but for all of life.

— ISAAC ASIMOV (1988)

Acknowledgements go here.

TABLE OF CONTENTS

Table of Contents	iii
List of Tables	ix
List of Figures	x
List of Abbreviations and Acronyms	xii
Abstract	xiii
Abstract	xv
Chapter 1: Introduction to the Ubiquitin 26S Proteasome System	1
1.1 The Ubiquitin 26S Proteasome System (UPS)	1
1.2 The 26S Proteasome	5
1.3 Proteasome Purification Strategies	12
1.4 26S Proteasome Substrate Processing	18
1.5 Proteasome Assembly	21
1.5.1 20S Core Protease Assembly	21
1.5.2 19S Regulatory Particle Assembly RP base	33
1.5.3 19S Regulatory Particle Assembly RP Lid	37

1.5.4	RP-CP Assembly and Alternative Capping Particles	39
1.5.5	Proteasome Assembly in Plants	39
1.6	Proteasome Isotypes	41
1.6.1	Immunoproteasome	41
1.6.2	Thymoproteasome proteasome	42
1.6.3	Testes-specific proteasome	43
1.6.4	$\alpha 4\text{-}\alpha 4$ proteasome Isotype	44
1.6.5	Plant Proteasome Isotypes	44
1.7	Conclusions	47
1.8	Literature Cited	50

Chapter 2: Morpheus Spectral Counter: A Computational Tool for Label-Free Quantitative Mass Spectrometry using the Morpheus Search Engine **65**

2.1	Summary	65
2.2	Main Text	66
2.3	Methods	80
2.3.1	Sample Preparation	80
2.3.2	Liquid Chromatography and High-Resolution Mass Spectrometry	82

2.3.3 Data Processing	83
2.3.4 Isoform Incorporation Rates	85
2.3.5 Speed and Accuracy Comparisons of Morpheus/MSpC to the TPP/ABACUS Pipeline	85
2.4 Discussion on Requiring Two Unique Peptides to Quantify a Protein	87
2.5 Tutorial	89
2.6 Acknowledgements	93
2.7 Literature Cited	93
Chapter 3: Proteomic Characterization of Affinity-Purified Arabidopsis Pro- teasome Particles Identifies Sub-Complex Specific Interactors	
3.1 Statement of Attribution	96
3.2 Summary	97
3.3 Introduction	99
3.4 Experimental Procedures	105
3.4.1 Transgenic Plants and Growth Conditions	105
3.4.2 Sequence and Phylogenetic Analyses	108
3.4.3 Proteasome Affinity Purifications using PAG1-FLAG and FLAG- RPT4a/b	109
3.4.4 SDS-PAGE and Native-PAGE Analysis and Silver Staining . .	110

3.4.5 MS/MS Sample Preparation	112
3.4.6 MS/MS Analyses	114
3.4.7 MS/MS Data Processing	115
3.4.8 Yeast-two hybrid, bimolecular fluorescent complementation, and HA-PAP1 immunopurification	118
3.5 Results	120
3.5.1 Arabidopsis Proteasomes Can Be Effectively Affinity-Purified Using Epitope-Tagged Versions of Either Isoform of the Reg- ulatory Particle Subunit RPT4	120
3.5.2 MS/MS Analyses of RPT4a/b Affinity-Purified Proteins Iden- tifies a Similar Complement of Proteins as PAG1-FLAG Affin- ity Purifications	125
3.5.3 Statistical Analysis of Affinity Enrichments for FLAG-RPT4a Show Little to No Difference When Compared to FLAG-RPT4b	127
3.5.4 Statistical Analyses of LFQ-MS/MS Data Show Affinity Prepa- rations Targeting Either the CP or RP Specifically Enrich for Their Respective Sub-complexes	129

3.5.5 Statistical Analysis of Affinity Purifications Enriched for the RP or CP Identifies Proteasome-Associated Proteins and Putative Assembly Chaperones Specific to Each Sub-complex	131
3.5.6 A Novel Plant Proteasome-Associated Protein, PAP1, Interacts with a Putative CP Assembly Chaperone	134
3.5.7 Inhibition of the Proteasome Results in Formation of Distinct Proteasome Sub-Species Containing Putative CP Assembly Chaperones	137
3.6 Discussion	141
3.7 Literature Cited	168
A: Affinity enrichment of the Post-Translational Modification Ubiquitin 177	
A.1 Summary	177
A.2 Introduction	177
A.3 Methods	177
A.4 Results	177
A.5 Discussion	177
A.6 Conclusions	177
A.7 Future Directions	177

LIST OF TABLES

2.1 Table of dNSAF values for each 20S proteasome subunit generated by analyzing the proteasome spike in experiments with the Morpheus and MSpC pipeline	78
2.2 Table of adj_NSAF values for each 20S proteasome subunit (equivalent to dNSAF) generated by analzying the proteasome spike in experiments with the TPP and ABACUS pipeline	79

LIST OF FIGURES

1.1 A schematic representation of the E1 activating, E2 conjugating, and E3 ligase enzymatic cascade for the covalent attachment of the small polypeptide Ubiquitin to target protein lysines	3
1.2 The 26S proteasome is made up of two functionally distinct sub-particles.	7
1.3 The 20S Core Protease (CP)	9
1.4 Cryo EM Structure of the 19S Regulatory Particle (RP)	12
1.5 Electron microscopy images of 20S and 26S proteasomes from mammals and plants	13
1.6 Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2)	19
1.7 Schematic representation of CP Assembly	24
1.8 Pba1 and Pba2 binding sites via their HbYX motifs into adjacent α -subunit pairs	30
1.9 Structural Diagram of opposing Pba1/2 Pba3/4 binding	32
1.10 Schematic representation of RP Assembly	36
1.11 Representation of subunits replaced for both the immunoproteasome and the thyoproteasome	46

1.12 Solvent accessible tag locations on the 26S proteasome for both PAG1 and RPT4	49
2.1 MSpC Graphic User Interface (GUI) and software flow chart	68
2.2 Confirmation of MSpC accuracy by analysis of MS/MS datasets gener- ated with the Universal Proteome Standard 2 (UPS2)	70
2.3 Re- Analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2)	72
2.4 Confirmation of MSpC accuracy by analysis of MS/MS datasets gen- erated with affinity purified <i>Arabidopsis</i> 20S proteasomes spiked into a total cell lysate from <i>E. coli</i> . Following MS/MS analysis, the dNSAF and uNSAF values for each subunit/isoform were determined by Morpheus and MSpC	75
2.5 MSpC combined with Morpheus works faster than TPP combined with ABACUS. Speed comparisons were performed for 12, 24, 48, and 96 raw MS/MS files generated with the 20S proteasome/ <i>E. coli</i> lysate samples analyzed in 2.4	81

LIST OF ABBREVIATIONS AND ACRONYMS

API	Application programming interface
BCA	Bicinchoninic acid protein assay
BLAST	Basic local alignment search tool
C#	C sharp, a programming language
Da	Dalton, the atomic mass unit
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
ESI	Electrospray ionization
E-value	Expectation value
FASTA	A format for storing protein sequences
FDR	False discovery rate
GUI	Graphical user interface
HCD	Higher-energy collisional dissociation
HPLC	High-performance liquid chromatography
LC	Liquid chromatography
<i>m</i>	Mass
min	Minute
MS	Mass spectrometry

MS ¹	Survey mass analysis
MS/MS	Tandem mass spectrometry
NCE	Normalized collision energy
nLC	Nanoflow liquid chromatography
ppm	Part per million
PSM	Peptide-spectrum match
PTM	Post-translational modification
s	Second
SILAC	Stable isotope labeling by amino acids in cell culture
S/N	Signal-to-noise ratio
TMT	Tandem mass tag

**ADVANCED PROTEOMIC CHARACTERIZATION OF THE 26S
PROTEASOME IN ARABIDOPSIS REVEALS INSIGHTS INTO
COMPOSITION AND ASSEMBLY**

David C. Gemperline

Under the supervision of Professor Richard D. Vierstra

At the University of Wisconsin-Madison

Abstract

The 26S proteasome is the central proteolytic effector in the ubiquitin system that is responsible for degrading numerous regulators following their selective ubiquitylation. While much is known about the construction of the yeast and mammal particles, little is known about the pathways used to assemble the plant particle. One challenge is that the known yeast chaperones appear sufficiently diverged to preclude high-confidence identification of their plant counterparts by genomic searches. Here, we used in-depth mass spectrometric analysis of Arabidopsis 26S proteasomes, which were affinity purified from seedlings under conditions that promote the accumulation of assembly intermediates, to identify a large collection of interacting proteins that associate with either the core protease (CP) or regulatory particle (RP). Sequence comparisons, Y2H and BiFC studies revealed

that some are likely assembly chaperones, with several CP factors harboring the signature C-terminal HbYX motif that allows their association with the α -subunit ring. Several of the RP-specific factors appear to be orthologs of the chaperones Nas2, Nas6, Hsm3 and Ecm29. Whereas yeast assembles only a single particle type, mammals can assemble alternate proteasomes by replacing individual subunits with distinct isoforms (e.g., immunoproteasomes). In plants, most 26S proteasome subunits are encoded by paralogous genes with sufficient divergence to suggest that plants also accumulate a collection of particles. However, proteomic analysis of proteasomes selectively enriched using paralog-specific tags strongly imply that although plants possess this genetic diversity, the incorporation of these paralogs appears random, and is mainly influenced by the differential expression of the corresponding genes. Taken together, these proteomic studies provide the first insights into plant proteasome assembly and diversity, and identify factors that build the CP and RP subcomplexes and finally the 26S holo-particle.

Richard D. Vierstra

ABSTRACT

Abstract

The 26S proteasome is the central proteolytic effector in the ubiquitin system that is responsible for degrading numerous regulators following their selective ubiquitylation. While much is known about the construction of the yeast and mammal particles, little is known about the pathways used to assemble the plant particle. One challenge is that the known yeast chaperones appear sufficiently diverged to preclude high-confidence identification of their plant counterparts by genomic searches. Here, we used in-depth mass spectrometric analysis of *Arabidopsis* 26S proteasomes, which were affinity purified from seedlings under conditions that promote the accumulation of assembly intermediates, to identify a large collection of interacting proteins that associate with either the core protease (CP) or regulatory particle (RP). Sequence comparisons, Y2H and BiFC studies revealed that some are likely assembly chaperones, with several CP factors harboring the signature C-terminal HbYX motif that allows their association with the α -subunit ring. Several of the RP-specific factors appear to be orthologs of the chaperones Nas2, Nas6, Hsm3 and Ecm29. Whereas yeast assembles only a single particle type, mammals can assemble alternate proteasomes by replacing individual subunits with distinct isoforms (e.g., immunoproteasomes). In plants, most 26S proteasome

subunits are encoded by paralogous genes with sufficient divergence to suggest that plants also accumulate a collection of particles. However, proteomic analysis of proteasomes selectively enriched using paralog-specific tags strongly imply that although plants possess this genetic diversity, the incorporation of these paralogs appears random, and is mainly influenced by the differential expression of the corresponding genes. Taken together, these proteomic studies provide the first insights into plant proteasome assembly and diversity, and identify factors that build the CP and RP subcomplexes and finally the 26S holo-particle.

Chapter 1

INTRODUCTION TO THE UBIQUITIN 26S PROTEASOME SYSTEM

Adapted in part from: R.S. Marshall*, D.C. Gemperline*, and R.D. Vierstra (2015). Purification of 26S proteasomes and their sub-complexes from plants. *Methods Mol. Biol.* DOI: 10.1007/978-1-4939-6533-5

*These authors contributed equally to this work.

This chapter has expanded sections on the proteasomes role in plant biology, proteasome assembly, and proteasome isotypes.

1.1. The Ubiquitin 26S Proteasome System (UPS)

Selective proteolysis in plants plays a critical role in both regulating growth and development, and maintaining cellular homeostasis (Nelson et al., 2014; Smallle and Vierstra, 2004; Vierstra, 1993, 2009). One of the principle pathways for protein degradation in plants and other eukaryotes is the ubiquitin-26S proteasome system (UPS), which involves the attachment of polyubiquitin chains to target proteins followed by their recognition and degradation by the 26S proteasome, an exquisitely designed proteolytic machine (Bhattacharyya et al., 2014; Finley, 2009; Vierstra,

2009). The UPS is highly conserved across all eukaryotes; it was first elucidated by elegant work in rabbit reticulocyte lysates (Ciechanover et al., 1980a,b; Etlinger and Goldberg, 1977; Hershko et al., 1980; Wilkinson et al., 1980), and was subsequently identified in other animals, yeast and plants (Ciechanover et al., 1984; Finley et al., 1984, 1987; Glotzer et al., 1991; Hochstrasser et al., 1991; Shanklin et al., 1987). Ubiquitin conjugation to target proteins is accomplished through a highly polymorphic, ATP-dependent cascade involving the sequential action of three enzyme classes, termed the E1 ubiquitin-activating enzymes, E2 ubiquitin-conjugating enzymes, and E3 ubiquitin-protein ligases (Berndsen and Wolberger, 2014; Smalle and Vierstra, 2004; Vierstra, 2009) (1.1). Selectivity in ubiquitylation is typically driven by the E3 family, which has dramatically expanded during plant evolution to include well over a thousand variants in *Arabidopsis* and other plant species (Hua et al., 2013; Hua and Vierstra, 2011). Indeed 6% of the *Arabidopsis* proteome is devoted to UPS (Smalle and Vierstra, 2004).

The UPS helps regulate plant growth and development by regulating several key proteins including those involved in morphogenesis, light sensing, and the circadian clock (reviewed in (Vierstra, 2009)). Some examples include E3 enzymes such as BIG BROTHER that controls organ size (Disch et al., 2006), F-BOX-LIKE 17 that helps regulate spermatogenesis (Kim et al., 2008), and UNUSUAL FLORAL

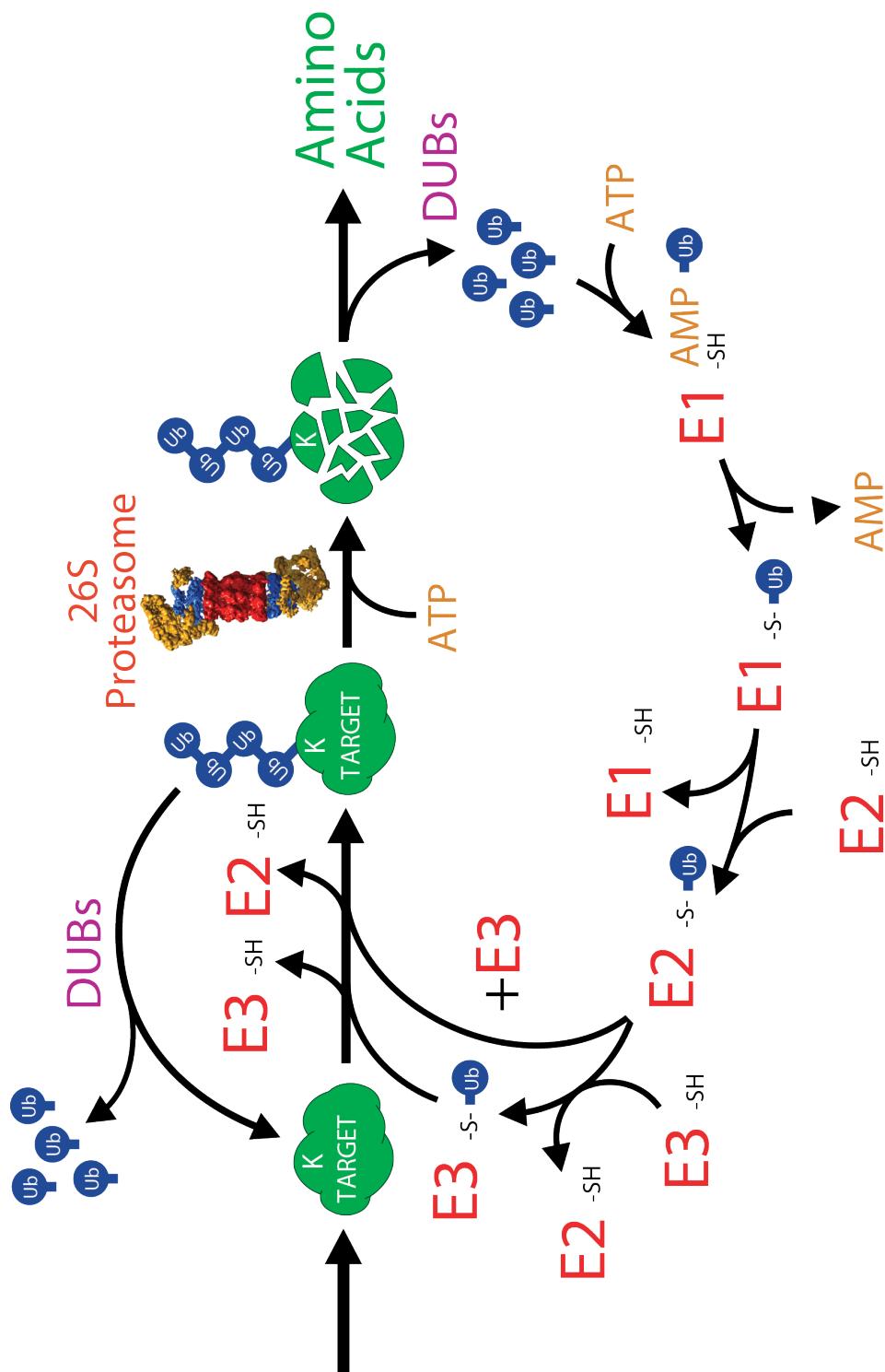


Figure 1.1: A schematic representation of the E1 activating, E2 conjugating, and E3 ligase enzymatic cascade for the covalent attachment of the small polypeptide Ubiquitin to target protein lysines. Once a protein becomes polyubiquitylated it is efficiently recognized by the 26S proteasome. The rate of addition of ubiquitins, and the rate of removal by de-ubiquitylating enzymes (DUBs) determines the final equilibrium of degradation.

ORGANS which regulates floral development (Samach et al., 1999). The UPS also helps regulate plants response to light by degrading phytochrome A (Clough and Vierstra, 1997; Shanklin et al., 1987), a key photoreceptor in plants. An E3 ligase ZTL also aides in removing TOC1 involved in circadian clock regulation (Más et al., 2003).

Additionally, the UPS also aides in the regulation of key plant hormonal responses including auxin, abscisic acid, jasmonic acid (JA), giberellins (GA), and ethylene. (reviewed in (Santner and Estelle, 2010)). Key hormone response factors are degraded by a variety of E3 complexes including TIR1 involved in auxin regulation (Dharmasiri et al., 2005), COI1, involved JA sensing (Katsir et al., 2008), and GID1 involved in GA responses (Murase et al., 2008). An elegant mechanism requires the direct binding of these key hormones either to their respective E3 complexes, or key proteins associated with these E3s to promote the efficient ubiquitylation and subsequent degradation of their hormone response factors (Shabek and Zheng, 2014).

The UPS is both co-opted and inhibited by several plant pathogens that helps aide in their virulence. The plant pathogen *Ralstonia solanacearum* encodes a suite of E3 enzymes that co-opt host machinery to evade plant defense responses (Angot et al., 2006). Additionally, *Pseudomonas syringae*, a tomato pathogen, secretes a small

molecule that irreversibly binds the host 26S proteasome to prevent plants from actively ubiquitylating and degrading key virulence factors (Schellenberg et al., 2010).

Self-incompatibility is a key process in plants that encourages plants to outcross helping aide in plant evolution and diversity by preventing inbreeding (Zhang et al., 2009). Upon incompatible pollination an E3 enzyme ARC1 helps to ubiquitylate compatibility factors in the pistil leading to self-pollen rejection (Stone et al., 2003). In another example S-RNase, which acts to degrade the RNA from self-pollen tubes preventing their growth, is actively ubiquitylated via SLF and SSK1 (McClure, 2004; Zhao et al., 2010). From these and other analyses it is clear that the UPS either directly, or indirectly, regulates many aspects of plant growth and development including hormone responses, pathogen interactions, and self-incompatibility. Therefore, a fundamental understanding of the UPS system in plants has important agricultural impacts.

1.2. The 26S Proteasome

The 26S proteasome is the central enzymatic effector of the UPS, and is a 2.5 MDa particle located in the cytosol and nucleus of eukaryotic cells. It is composed of two functionally distinct sub-complexes; the 20S core protease (CP) that houses

the proteolytic active sites, and the 19S regulatory particle (RP) that recognizes appropriate substrates (1.2 A and B; (Bhattacharyya et al., 2014; Finley, 2009; Lander et al., 2012; Lasker et al., 2012; Unverdorben et al., 2014)).

The CP has a barrel shape generated by four stacked hetero-heptameric rings, which contain seven α -subunits or seven β -subunits (termed PAA-PAG and PBA-PBG, respectively, in *Arabidopsis*) in an α 1-7/ β 1-7/ β 1-7/ α 1-7 configuration (1.3 A (side view- stacked rings), C (top view- α -subunits) and D (top view- β -subunits).

Upon assembly, a central chamber is formed at the β -ring interface that houses six peptidase catalytic sites provided by the β 1 (PBA), β 2 (PBB), and β 5 (PBE) subunits (Arendt and Hochstrasser, 1997; Heinemeyer et al., 1997) (1.3 B and E). The active sites involve a catalytic triad, one residue of which is an N-terminal threonine that becomes exposed during CP assembly. Collectively these peptidases can cleave a broad range of protein sequences with peptidylglutamyl-peptide hydrolyzing (PGPH) (β 1), trypsin-like (β 2), and chymotryptic-like (β 5) activities (Arendt and Hochstrasser, 1997; Groll et al., 1999) (1.3 E). The α -rings create two antechambers with narrow opposing axial pores that are gated by extensions at the N-terminus of several subunits (1.3 B) (Groll et al., 2000; Ruschak et al., 2010). Through this distinctive architecture, the CP acts as a self-compartmentalized protease that will only degrade polypeptides that are deliberately recognized, unfolded, and imported

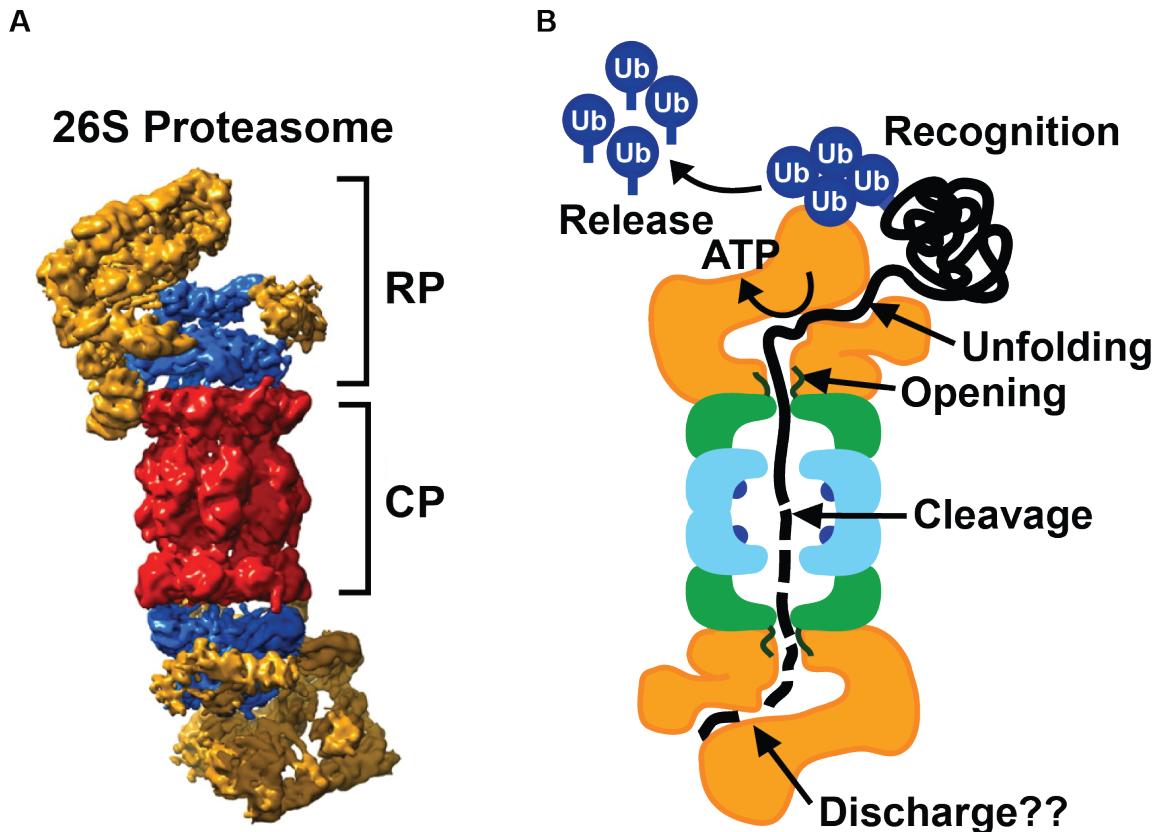
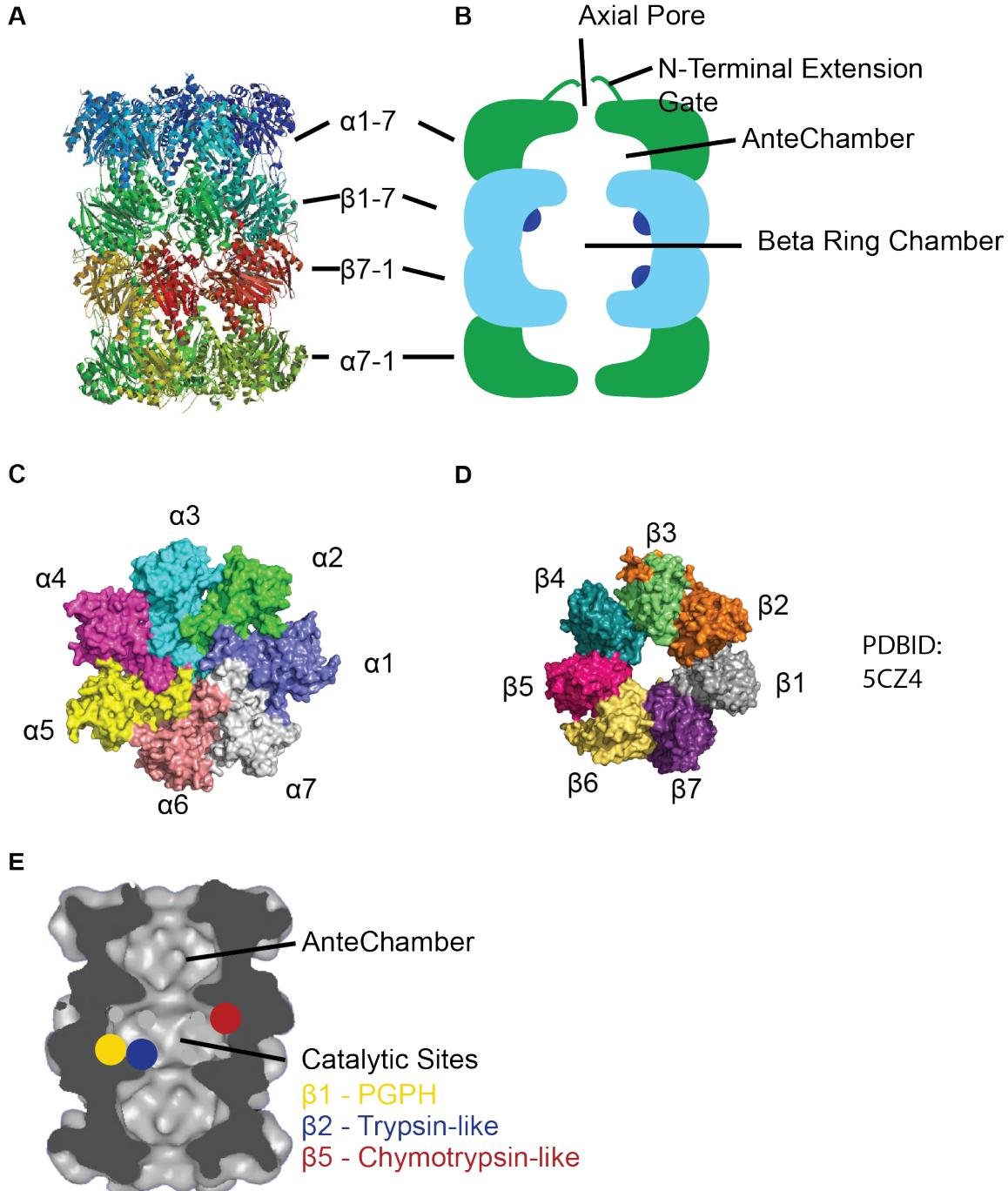


Figure 1.2: The 26S proteasome is made up of two functionally distinct sub-particles.. (A) The Core Protease (CP) is shown in red, while the Regulatory Particle (RP) is shown in blue for the RPT subunits, and in gold for the RPN sub-units (adapted from (Lasker et al., 2012)). (B) Schematic representation of CP and RP functions. The RP is responsible for recognizing and unfolding polyubiquitylated substrates into a central axial chamber upon which the unfolded polypeptide is cleaved by the CP. The RP is also responsible for releasing the ubiquitin so that it can be recycled for re-use through the de-ubiquitylating activity of the metalloprotease RP subunit RPN11.

[Figure 1.3 caption follows on next page]



into the β -ring chamber.

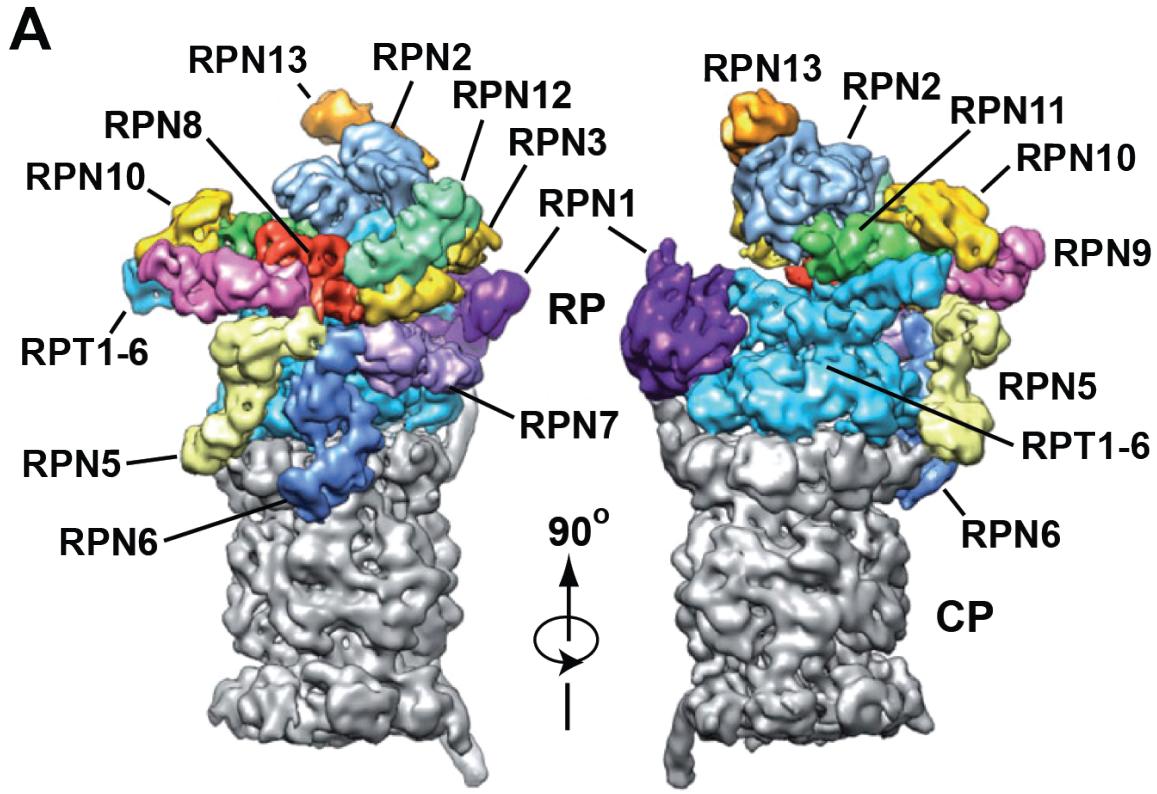
The CP is capped at one or both ends by the RP, which sits on top of the axial pores. The RP provides activities for recognition of ubiquitylated proteins, substrate unfolding and import, and release of the ubiquitin moieties before substrate degradation (1.2). Its binding to the CP is stabilized by ATP, which is thus a necessary ingredient for purifying intact 26S proteasomes (Smith et al., 2005). The RP itself consists of two sub-complexes; the base, which contains a hexameric ring of AAA-ATPases (RPT1-6) plus two non-ATPase subunits, RPN1 and RPN2; and the lid, which is composed of an additional 11 non-ATPase subunits, RPN3, RPN5-13 and DSS1/SEM1 (1.2; (Bhattacharyya et al., 2014; Book et al., 2010; Finley, 2009;

Figure 1.3 (preceding page): The 20S Core Protease (CP). (A) An x-ray crystallographic structure of the 20S CP shows its barrel-like structure. The structure is composed of four stacked hetero-heptameric rings in an α 1-7, β 1-7, β 1-7, α ,1-7 configuration. (B) This stacked configuration generates an interior structure amenable to cleaving unfolded polypeptides that enter through the axial pore. N-terminal extensions present in α -subunits gate the entry to the antechamber. The antechamber helps maintain the polypeptide in an unfolded state, and the β -ring chamber houses the catalytic activities of the CP. (C) Top-down view of the CP showing the α -subunit configuration (PDB ID 5CZ4) visualized in PyMOL (Schrodinger, 2015). (D) Top-down view of the CP showing the β -subunit configuration (PDB ID 5CZ4) visualized in PyMOL. (E) Catalytic activities of the 20S proteasome are conferred by the β 1 (peptidylglutamyl-peptide hydrolyzing (PGPH)), β 2 (Trypsin-like), and β 5 (Chymotrypsin-like) subunits. A slice through the middle of a model of the 20S proteasome (adapted from (Kisselev et al., 2012)) shows the location and type of the three catalytic activities of the 20S proteasome. While this is a half-structure, each barrel actually houses 6 catalytically active locations, with three different catalytic activities.

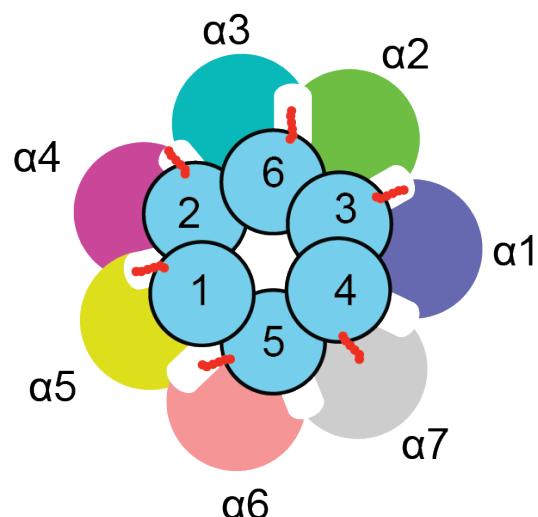
Glickman et al., 1998b; Russell et al., 2013). This lid/base demarcation was first revealed by the absence of lid subunits in proteasomes isolated from a *rpn10* yeast deletion strain, and it was hence thought that RPN10 helps enforce binding of the lid to the base (Glickman et al., 1998a). However, more recent structural studies have demonstrated that RPN10 has a more indirect stabilizing role via its interaction with RPN9 (Lander et al., 2012). The ring of RPT subunits in the base promotes substrate unfolding through ATP hydrolysis, and gates the α -ring axial pores through repositioning of the CP α -subunit extensions (Köhler et al., 2001; Rabl et al., 2008; Smith et al., 2005). The N-terminal regions from proximal RPT pairs intertwine to create three spokes onto which most RPN subunits are scaffolded (1.4 Å; (Beck et al., 2012)).

Several RPT subunits harbor C-terminal *Hydrophobic-Tyrosine-Any* amino acid (HbYX) motifs important for their binding to the α -subunits in the CP. These C-terminal HbYX and HbYX-like extensions typically intercalate between adjacent α -subunits (1.4 B). The RPN6 subunit acts as a molecular clamp to tether the RP onto the CP (1.4 A) (Pathare et al., 2012).

[Figure 1.4 caption follows on next page]



B



→ HbYX /
HbYX-like
C-terminal Extensions

1.3. Proteasome Purification Strategies

Even before the realization that the 26S proteasome is a protease, sub-particles of the complexes were described. The first reports of proteasomes used avian erythroblast preparations enriched by differential ultracentrifugation followed by fractionation through a sucrose gradient (Schmid et al., 1984). These 20S fractions isolated in the absence of added ATP were found to inhibit mRNA translation in a cell-free system, leading to early proposals that the identified complex repressed gene expression through a cryptic ribonuclease activity. This lead to the particle initially being named the prosome (Kremp et al., 1986; Schmid et al., 1984). Subsequent analyses of these preparations by SDS-PAGE and electron microscopy revealed the signature ladder of α - and β -subunits at 20-35 kDa, as well as their barrel-like architecture (1.5 Å and B; (Baumeister et al., 1988; Kremp et al., 1986; Schmid et al., 1984)).

Purification of the 20S fraction from HeLa cells followed by SDS-PAGE also gave rise to this stereotypical protein banding pattern and shape (Schmid et al., 1984),

Figure 1.4 (preceding page): Cryo EM Structure of the 19S Regulatory Particle (RP). (A) The RP structure as solved by Cryo-EM is shown (modified from (Lander et al., 2012)). In the center in blue are the RPT1-6 subunits, while the RPN subunits are individually labeled. A 90-degree rotation is shown so that the other side of the RP can be visualized. (B) A schematic representation of how the RP subunit C-termini containing HbYX and HbYX-like motifs intercalate between the α -subunits (adapted from (Sokolova et al., 2015)). Gaps between each α -subunit are shown, and the HbYX motifs are shown as red strands.

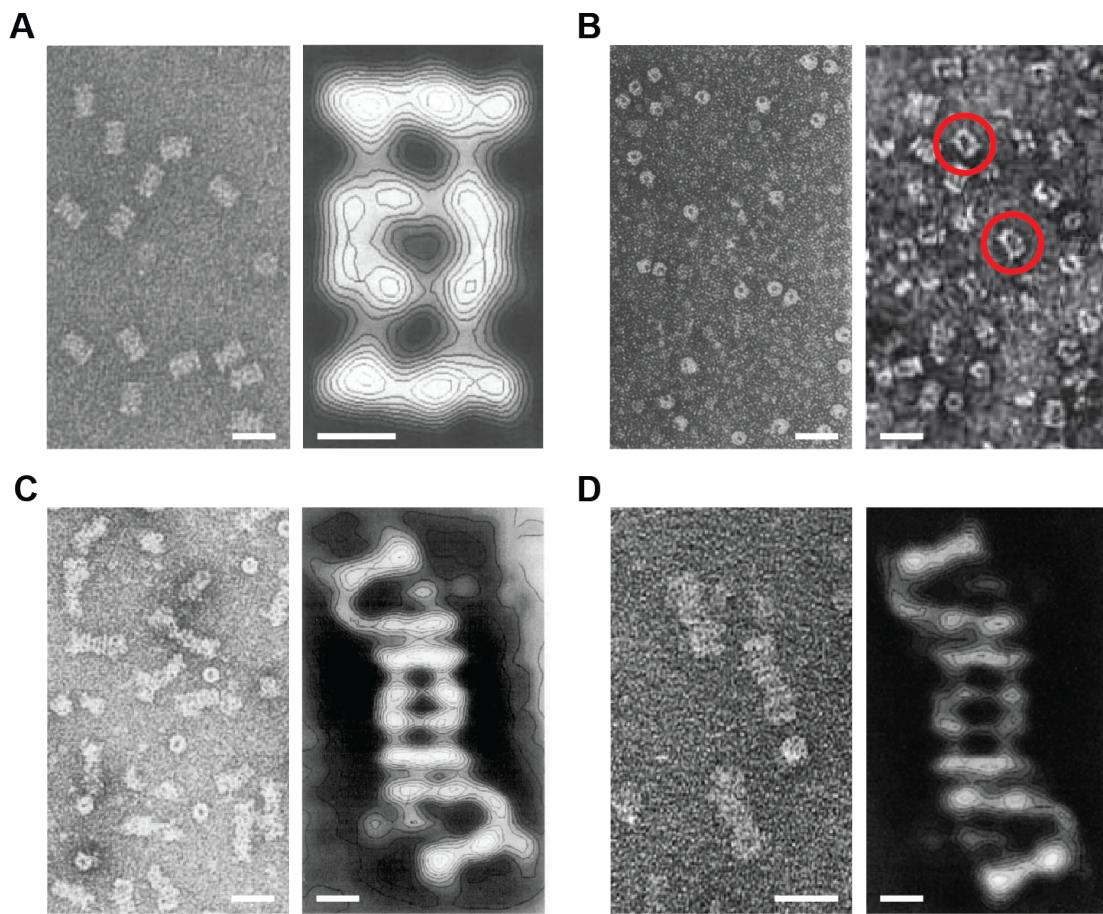


Figure 1.5: Electron microscopy images of 20S and 26S proteasomes from mammals and plants. (A) Images of 20S proteasomes purified from rat skeletal muscle. On the left is an electron micrograph of the 20S particles negatively stained with sodium phosphotungstate, while on the right is a close-up image with overlaid contour plots generated by correlation averaging of approximately 300 individual images negatively stained with ammonium molybdate. (B) Images of the first 20S proteasomes purified from different plant species. On the left are proteasomes isolated from tobacco leaves, while on the right are proteasomes from potato tubers, both negatively stained with uranyl acetate. The typical barrel-shaped structures are indicated with red circles. (C) Images of 26S proteasomes purified from rat liver. On the left is an electron micrograph of the 26S particles negatively stained with uranyl acetate, while on the right is a close-up image with overlaid contour plots generated by correlation averaging of 215 individual images. (D) Images of 26S proteasomes purified from spinach leaves. On the left is an electron micrograph of the 26S particles negatively stained with uranyl acetate, while on the right is a close-up image with overlaid contour plots generated by correlation averaging of 450 individual images. In all cases, scale bars represent 25 nm for the electron micrograph images and 5 nm for close-up images generated by averaging. The images were modified from references (Baumeister et al., 1988; Fujinami et al., 1994; Kremp et al., 1986; Schliephacke et al., 1991; Yoshimura et al., 1993).

and this was followed shortly thereafter by the first description of plant prosomes, purified from tobacco leaf extracts using similar sedimentation protocols in ATP-free buffers (Kremp et al., 1986). In these later cases, the purified preparations had strong peptidase activity but little to no RNase activity, thus leading to the conclusion that the CP is actually a protease. Once its true function in protein turnover was confirmed, the moniker for the particle was changed to proteasome (Arrigo et al., 1988).

Subsequently, the 20S particle was purified from other plant tissues, including dry pea seeds, potato tubers, mung bean seedlings, and leaves from both spinach and wheat (Murray et al., 1997; Ozaki et al., 1992; Schliephacke et al., 1991; Skoda and Malek, 1992). These purifications were typically performed using sequential anion exchange and size-exclusion chromatography steps in the absence of ATP, hence only the CP was isolated. Their remarkable similarity in protein composition and structure, as observed by SDS-PAGE and electron microscopy, respectively, coupled with the fact that several of the plant subunits cross-reacted with antibodies against their yeast, human, rat and *Xenopus* counterparts, strongly implied that the CP was conserved and widely distributed among eukaryotes (Schliephacke et al., 1991).

The complete 26S proteasome (i.e. the CP capped at one or both ends by the RP)

was subsequently discovered by the purification of ubiquitin conjugate-degrading activity from rabbit reticulocytes (Hough et al., 1986). While it had been well established that major catabolic processes in animal cells involved the ATP-dependent proteolysis of selective substrates (Etlinger and Goldberg, 1977), the enzyme(s) responsible for this activity had yet to be identified. Taking advantage of the new ability to synthesize ubiquitylated substrates such as ^{125}I -labelled ubiquitin-lysozyme conjugates (Hough and Rechsteiner, 1986), a protocol was developed to purify the responsible ATP-dependent protease. Through a series of anion exchange and size exclusion chromatography steps followed by glycerol gradient sedimentation, all of which were performed in ATP-containing buffers, the responsible activity was isolated (Hough et al., 1986, 1987). The active enzyme turned out to be the 20S proteasome (i.e. the CP) along with a number of additional polypeptides, which together formed a 26S particle, thus providing the first direct link between ubiquitylation and a protease (Ganoth et al., 1988; Hough et al., 1987; Waxman et al., 1987). SDS-PAGE analysis of these preparations identified a host of new polypeptides in the 35-100 kDa range in addition to the known CP subunits, which were later shown to comprise a second stable complex, the RP. Shortly thereafter, the RP was demonstrated to have ATPase activities attributable to the RPT subunits, which help in substrate unfolding and maintaining CP-RP association (Armon

et al., 1990). Electron microscopic images of the full 26S particle then revealed its diagnostic quaternary structure in which the CP is capped by one or two RPs that sit over the axial pores for substrate entry (1.5 C; (Peters et al., 1991; Yoshimura et al., 1993)).

The existence of a similar 26S proteasome in plants was initially implied by the detection of an ATP-dependent activity in oat and wheat germ extracts capable of degrading ubiquitylated proteins (Hatfield and Vierstra, 1989; Vierstra and Sullivan, 1988). This was followed some years later by the first isolation of a complete plant 26S proteasome holocomplex from spinach leaves (Fujinami et al., 1994). As with the mammalian forms, purification was achieved by anion exchange and size exclusion chromatography, followed by glycerol gradient centrifugation, all in the presence of ATP to stabilize the CP-RP association. These spinach preparations were, like their rabbit reticulocyte counterparts, able to rapidly degrade ubiquitylated substrates in an ATP-dependent manner, and further analysis by native-PAGE, SDS-PAGE and electron microscopy revealed the complete subunit composition and caterpillar-like structure of the plant particle (1.5 D; (Fujinami et al., 1994)). Similar purifications were successful using rice suspension culture cells and garlic cloves (Malik et al., 2004; Yanagawa et al., 1999), which were accompanied by the first demonstrations that proteasome inhibitors designed for their mammalian counterparts were effec-

tive with the plant particles, suggesting very similar enzymatic mechanisms (Ozaki et al., 1992; Woffenden et al., 1998).

Despite its prevalence as a genetic model, purification of the 26S proteasome from the flowering plant *Arabidopsis thaliana* was not reported until several years after other plant species (Yang et al., 2004). First protocols involved differential PEG precipitation followed by anion exchange and size exclusion chromatography, with the latter exploiting the large size of the holoprotease. More recently, an improved one-step affinity method was developed (Book et al., 2010), based on the strategies that had been successfully employed in yeast (Leggett et al., 2005). Here, epitope-tagged proteasomes were generated by genetically replacing the subunit PAG1 with a variant bearing a C-terminal FLAG tag; this tagged particle could then be purified with appropriate affinity matrices, and released in non-denaturing conditions with FLAG peptide. This approach enables rapid and robust purification of the whole 26S proteasome complex when performed in the presence of ATP, or enables purification of the CP sub-particle when performed in the absence of ATP (Book et al., 2010). This affinity method has considerable advantages compared to previous, conventional chromatographic approaches (Yang et al., 2004) as it is both faster, more reliable and produces higher yields per gram of tissue (~6 µg/g). This milder more rapid technique also prevents breakdown of some subunits, in particular RPN10, which

is sensitive to post-homogenization proteolysis (Yang et al., 2004). One caveat is that the epitope tag, given its exposed position and flexible structure, might be sensitive to proteolytic cleavage following tissue homogenization. For the PAG1-FLAG protocol, chymostatin was found to effectively block the interfering protease (Book et al., 2010). An example of such preparations analyzed by SDS-PAGE followed by immunoblotting with antibodies against several proteasome subunits, is shown in 1.6.

1.4. 26S Proteasome Substrate Processing

A variety of proteins help the 26S proteasome process ubiquitylated substrates. Some include key constituents of the complex itself, such as RPN11, which is a metalloprotease that uses a zinc-coordinated active site to release the ubiquitin moieties isopeptide-linked to substrates (Verma et al., 2002; Worden et al., 2014). Through RPN11 and other loosely associated deubiquitylating enzymes such as UBP6/USP14 (Hanna et al., 2006; Sakata et al., 2011), bound ubiquitins are actively recycled. Substrate selection by the 26S proteasome is dictated by several ubiquitin receptors intrinsic to the RP lid, including RPN1, RPN10, RPN13, and DSS1/SEM1 (Elsasser et al., 2002; Fatimababy et al., 2010; Finley, 2009; Lin et al., 2011; Paraskevopoulos et al., 2014; Sakata et al., 2012; Shi et al., 2016; Van Nocker

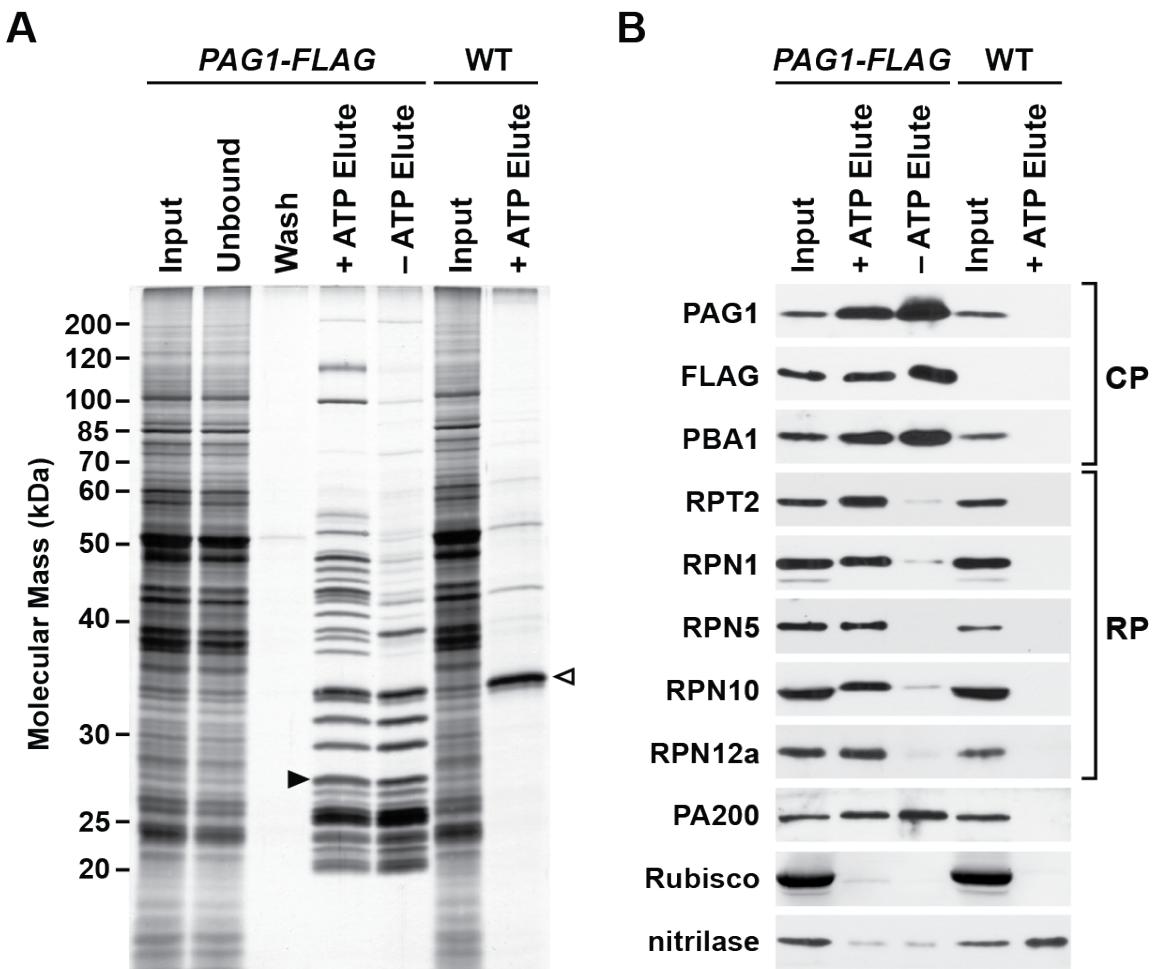


Figure 1.6: Confirmation of MSpc accuracy by analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2). Affinity purification of 26S proteasomes from *PAG1::PAG1-FLAG pag1-1* plants

(Book et al., 2010) **(A)** SDS-PAGE analysis of proteins obtained at each affinity purification step. Total protein extracts from 10 day old wild-type (WT) and *PAG1::PAG1-FLAG pag1-1* plants were incubated with anti-FLAG affinity resin, washed, and competitively eluted with the FLAG peptide. The procedure was performed in the presence or absence of ATP, and the input, unbound, washed and eluted fractions were subjected to SDS-PAGE and the gel stained for protein with silver. The black arrowhead indicates the *PAG1-FLAG* protein, while the open arrowhead identifies nitrilase, which is non-specifically enriched during the purification. **(B)** Immunoblot detection of various 26S proteasome subunits in the affinity-purified preparations shown in A. Subunits tested include the CP subunits PAG1 and PBA1, the RP subunits RPT2, RPN1, RPN5, RPN10 and RPN12a, and the alternate capping particle PA200. Other proteins tested include the Rubisco small subunit and nitrilase. This figure was modified from reference (Book et al., 2010).

et al., 1996). RPN10 binds ubiquitin via defined ubiquitin-interacting motifs (UIMs), of which yeast, human and *Arabidopsis* RPN10 contain 1, 2 and 3 in tandem, respectively (Fatimababy et al., 2010; Finley, 2009; Fu et al., 1998; Lin et al., 2011; Van Nocker et al., 1996). By contrast, RPN13 binds ubiquitin via a pleckstrin-like receptor for ubiquitin (PRU) domain, which is structurally distinct from UIMs but binds to the same hydrophobic patch on ubiquitin (Husnjak et al., 2008; Schreiner et al., 2008). More recently, DSS1/SEM1 was also found to be a proteasomal ubiquitin receptor (Paraskevopoulos et al., 2014). It had previously resisted identification due to both its small size, which prevented visualization by standard protein stains following sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and its paucity of lysine and arginine residues, which complicated detection by conventional mass spectrometric methods. Only with the use of top-down mass spectrometry of 26S proteasome complexes was DSS1/SEM1 first detected in intact 26S proteasomes from *Arabidopsis* (Russell et al., 2013). In addition to these core ubiquitin receptors, there are several extra-proteasomal ubiquitin-binding proteins that shuttle ubiquitylated cargo to the RP. They work by virtue of ubiquitin-associated (UBA) domains that bind ubiquitin, combined with a ubiquitin-like (UBL) domain that interacts with the intrinsic ubiquitin receptors such as RPN10. Important shuttle factors in plants include RAD23, DSK2, and DDI1 (Farmer et al.,

2010; Fatimababy et al., 2010; Finley, 2009; Lin et al., 2011), though many other ubiquitin-binding proteins are known in other species (Husnjak and Dikic, 2012). Numerous other factors also associate sub-stoichiometrically with the mature CP and RP sub-complexes, including deubiquitylating enzymes, several E3 ligases and protein kinases, and a collection of protein folding chaperones (Besche et al., 2014; Book et al., 2010; Leggett et al., 2002; Xie and Varshavsky, 2000).

1.5. Proteasome Assembly

Not surprisingly given its intricate architecture, construction of the eukaryotic 26S proteasome requires a large collection of assembly factors that work in synchrony. The most well-studied eukaryotic models for assembly include both yeast and mammalian systems; however, very little if anything is known about the assembly of the plant complex.

1.5.1. 20S Core Protease Assembly

In yeast and mammals five of the seven β -subunits ($\beta 1$, $\beta 2$, $\beta 5$, $\beta 6$, and $\beta 7$) are synthesized as an immature propeptide form with an N-terminal extension. These β -subunits undergo autolytic cleavage exposing a catalytically active N-terminal threonine for the $\beta 1$, $\beta 2$, $\beta 5$ subunits upon CP maturation. (Gu and Enenkel,

2014; Lee et al., 1990). This propeptide prevents inactivation of the mature forms N-terminal threonine by N-terminal acetylation, and also prevents premature activation of peptidase activities (Arendt, 1999; Arendt and Hochstrasser, 1997). While this propeptide processing occurs in *Arabidopsis*, and likely other plants, its role in plant proteasome assembly remains to be investigated (Book et al., 2010). In yeast, complexes containing these immature propeptide forms were first identified in 13-16S fractions (smaller than the 20S fraction) suggesting that distinct assembly steps may occur (Frentzel et al., 1994). Indeed in yeast these propeptides were found to be important in the assembly of the 20S CP as deletion of the $\beta 5$ propeptide was shown to be lethal by preventing assembly of the CP, and deletion of the $\beta 2$ propeptide accumulated aberrant complexes (Jager et al., 1999). Several of these propeptides were found to interact with the first dedicated assembly chaperone discovered, Ump1, which was identified in a screen for null mutants defective in ubiquitin-mediated proteolysis in yeast (Li et al., 2007; Ramos et al., 1998). Ump1 along with other dedicated assembly chaperones aide in the assembly of 15S complexes, or proteasome half barrels containing α -subunits 1-7, and β -subunits 1-6, noticeably lacking the $\beta 7$ subunit (Li et al., 2007; Marques et al., 2007). Similar investigations of the mammalian $\beta 5$ propeptide enabled identification of 15S half-barrels that also contained Ump1 and a similar complement of α - and β -subunits also lacking

the $\beta 7$ subunit (Witt et al., 2000). Later studies would go on to show that in both mammals and yeast, $\beta 7$ is the last subunit to enter the β -ring, resulting in Ump1 being degraded as the first substrate of nascent CP (Hirano et al., 2008; Ramos et al., 1998). Intriguingly, the alternative capping particle, PA200 co-purified with these 15S half-barrels suggesting that PA200 may be involved in proteasome assembly (see CP assembly summary in 1.7) (Li et al., 2007; Marques et al., 2007).

An additional suite of assembly chaperones aide in construction of the CP including the proteasome biogenesis associated Pba1-4 proteins in yeast, and the proteasome assembly chaperones PAC1-4 in mammals. The yeast assembly chaperones Pba1 and Pba2 were initially discovered in affinity preparations of the CP in a genetic background that induced the formation of aberrant 15S half-barrels (Li et al., 2007). MS/MS analyses of these 15S complexes separated by native-PAGE identified Pba1, Pba2, and Ump1 contained in these complexes suggesting that Pba1 and Pba2 played a role in CP assembly (Li et al., 2007). In a separate study, Pba1 and Pba2, along with two additional assembly chaperones Pba3 and Pba4 were identified in a mutant screen for suppressors of a conditional dominant lethal allele of Rad53 which induces DNA damage (Le Tallec et al., 2007). Pba1, and Pba2 were found to co-fractionate with both the precursor propeptide form of the $\beta 2$ subunit, and the mature form; however, Pba3 and Pba4 were found to exclusively fractionate

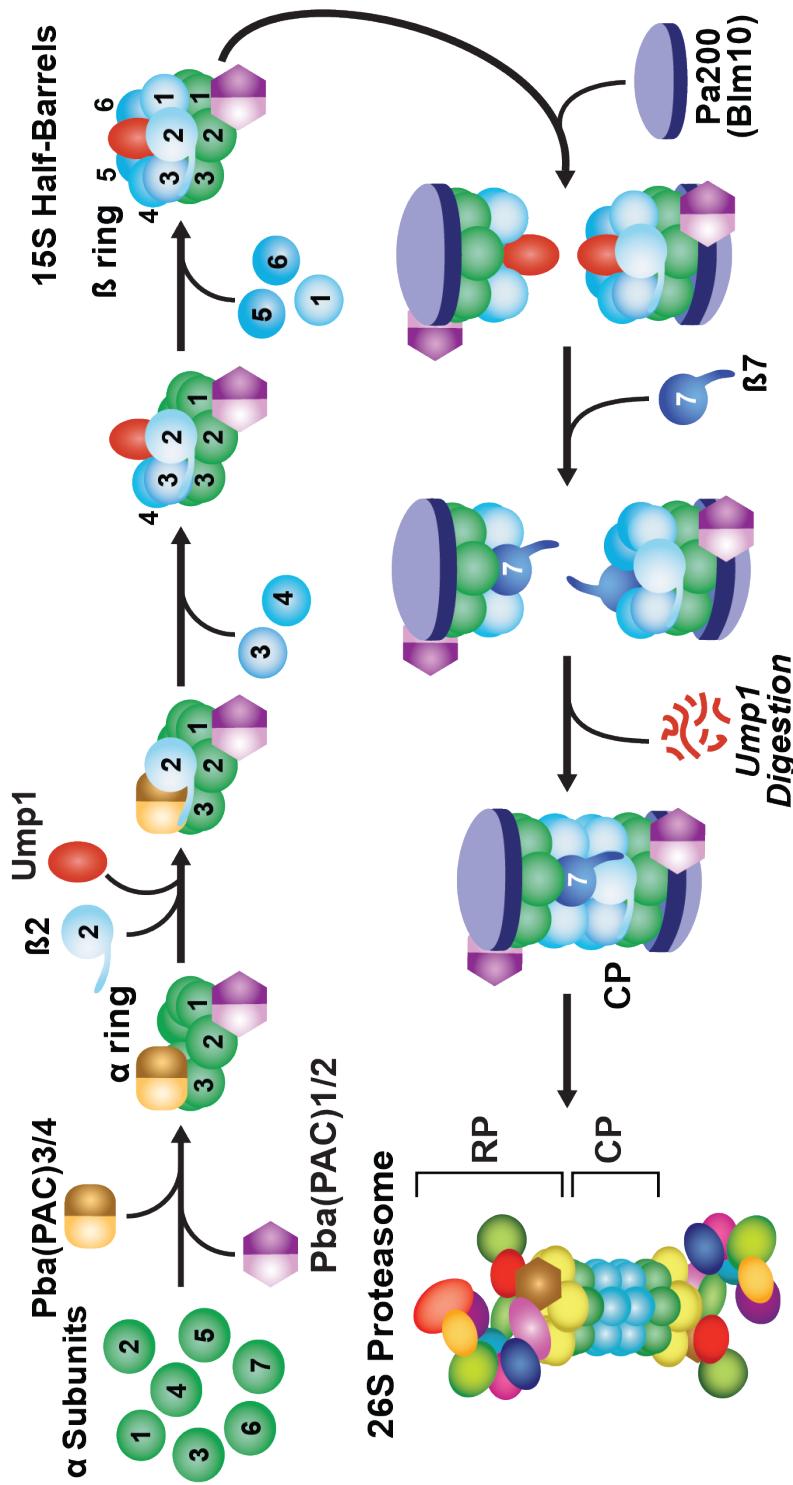


Figure 1.7: Schematic representation of CP Assembly. Adapted from (Hirano et al., 2008). Pba1-4 (yeast) and PAC1-4 (human) nomenclature is used. Free α -subunits are bound to prevent aberrant α -ring assembly by Pba1-4. The β 2 subunit along with Ump1 bind this α -ring, and subsequently the Pba(PAC)3/4 heterodimer leaves the assembly complex. Then the β 3 and β 4 subunits are incorporated and the β 5 and β 6 subunits are added forming a 15S half-barrel, possibly with the help of PA200. Two half-barrels come together and the last β -subunit, β 7 is added. N-terminal propeptides for β -subunits are autocatalytically cleaved, Ump1 is degraded as the first substrate of the nascent proteasome, and a fully functional CP is generated.

with the propeptide form of β 2 (Le Tallec et al., 2007). These data suggested that Pba3 and Pba4 aided in earlier stages of assembly, while Pba1 and Pba2 seemed to act at both early and later stages. Pba1 and Pba2 were found to form a heterodimer, while Pba3 and Pba4 were found to form a separate heterodimer (Le Tallec et al., 2007).

An orthologous system to the yeast Pba1-4 system was elucidated in mammals. PAC1 and PAC2 were first identified in affinity preparations of an immature propeptide form of the β 1 subunit, and like their yeast counterparts (Pba1 and 2) were found to form a stable heterodimer (Hirano et al., 2005). Consistent with their role in assembly they were found to co-fractionate with immature β -subunits, and Ump1 (Hirano et al., 2005). Overexpression of these PAC1 and PAC2 chaperones yielded increases in half-barrel precursors, while knockdown by siRNA impaired assembly of the CP complex (Hirano et al., 2005). An additional chaperone involved in assembling the α -ring was identified in PAC1 immunopurifications in human cell lines, and was termed PAC3 (Hirano et al., 2006). siRNA knockdown of PAC3 attenuated α -ring formation and gave rise to the accumulation of free α -subunits, suggesting that PAC3 played a role in stabilization of the CP α -ring (Hirano et al., 2006). Mammalian PAC4 was identified as an ortholog to yeast Pba4 through PSI-BLAST analyses (Kusmierczyk et al., 2008) and together with PAC4

forms a heterodimer, like its yeast Pba3/4 counterpart (Le Tallec et al., 2007).

The role of the PAC1/2 and PAC3/4 heterodimers in assembly, along with the exact ordered incorporation of β -subunits was aided by siRNA knockdown of specific β -subunits, and studying the resulting assembly products (Hirano et al., 2008). Through this analysis Hirano *et al.* clearly delineated PAC3/4's role in early assembly, and found that PAC3 and PAC4 must leave the assembling half-barrel before the addition of the β_3 subunit. While this was demonstrated in mammals, observations of the yeast Pba3/4 heterodimer suggest a similar model in which the Pba3/4 heterodimer also remains bound to the α -ring until the β_2 subunit binds (Kusmierczyk et al., 2008; Yashiroda et al., 2008). While Pba(PAC)3/4 act early in assembly, Pba(PAC)1/2 have the propensity to bind the mature proteasome, and stay bound even after the incorporation of β -subunits.

A model of yeast and mammalian assembly adapted from Hirano *et al.* is as follows (1.7). After the α -ring is formed, with the help of Pba(PAC)1/2 and Pba(PAC)3/4, the β_2 subunit is incorporated along with its C-terminal tail that helps to recruit the β_3 subunit (Hirano et al., 2008). Pba(PAC)3/4 is released upon β_3 subunit binding (Hirano et al., 2008). Ump1 stabilizes the incorporation of additional β -subunits by binding their inherently disordered propeptide regions (Li et al., 2007). β_4 , β_5 , β_6 , and β_1 are then incorporated (Hirano et al., 2008).

At this stage PA200 may help stabilize this precursor; however, its exact role in assembly is poorly understood. Two 15S half proteasomes come together, and finally the β_7 subunit is incorporated (Hirano et al., 2008). The β -subunit propeptides are then cleaved exposing the catalytically active N-terminal threonines and Ump1 is degraded becoming the first substrate of the nascent CP (Li et al., 2007). Pba(PAC)1/2 are then released, or can bind mature proteasomes until activation by the RP (Hirano et al., 2008). Taken together these data are consistent with a model in which the heterodimeric pair Pba(PAC)3/4 acts at early stages to assemble free α -subunits into an α -ring while preventing the formation of aberrant α -subunit complexes (Hirano et al., 2008). In contrast, the heterodimeric pair Pba(PAC)1/2 acts at both early and later stages in proteasome assembly and has the ability to bind mature CP (Wani et al., 2016).

While the CP assembly process is conserved in yeast and mammals, the orthologs involved share very little sequence homology. The Ump1 orthologs share ~20% identity, and Pba1-4 and PAC1-4 share ~10% identity respectively (Murata et al., 2009). Despite this low sequence identity, structural analysis revealed that the Pba3/4 heterodimer shared significant structural features with the human PAC3 protein, suggesting that these assembly chaperones may have retained structural similarities (Yashiroda et al., 2008). The Hochstrasser group also speculated that

the Pba1/2 and PAC1/2 complexes may be structurally related (Kusmierczyk et al., 2011). Besides these speculative structural implications, Pba1 and Pba2 share some protein sequence features, including HbYX motifs important for their interaction with the α -ring (Kusmierczyk et al., 2011). These features are conserved in their mammalian orthologs with PAC1 containing a HbYX motif, while PAC2 shares a HbYX-like HbY/F motif that is conserved in other eukaryotes outside the fungal kingdom (Kusmierczyk et al., 2011).

Structural analysis of these CP assembly chaperones have also provided insight into their function. Pba1, and Pba2 were able to bind mature 20S proteasomes, and a co-crystal structure has been solved by X-ray crystallography (Stadtmueller et al., 2012). The HbYX motif found in Pba1 was shown to intercalate between the α 5 and α 6 subunits, while the HbYX motif found in Pba2 was shown to intercalate between the α 6 and α 7 subunits (1.8). Crystal structures of the Pba3, Pba4 complex were shown to be similar to the PAC3 suggesting similar role (Yashiroda et al., 2008). A co-crystal structure binding the α 5 subunit helped define the binding surface of the Pba3/4 heterodimer (Yashiroda et al., 2008). This binding surface for Pba3/4 occurs on the opposite side of the α -ring as compared to the Pba1/2 binding surface providing further structural evidence for the distinct and separate roles that the Pba1/2 and Pba3/4 play in proteasome assembly (see 1.9) (Stadtmueller et al., 2012;

Yashiroda et al., 2008).

Outside of eukaryotes, proteasome assembly has also been studied in *archeal* and *eubacterial* proteasome models (Kusmierczyk et al., 2011). While the assembly of these proteasomes can occur without assembly chaperones two chaperones, PbaA and PbaB were identified that likely aide in efficient assembly of the particle (Kusmierczyk et al., 2011). PbaA and PbaB are evolutionarily related to yeast Pba1 and Pba2 and both harbor conserved HbYX motifs showing that these chaperones provide an ancient link to proteasome assembly evolution (Kusmierczyk et al., 2011).

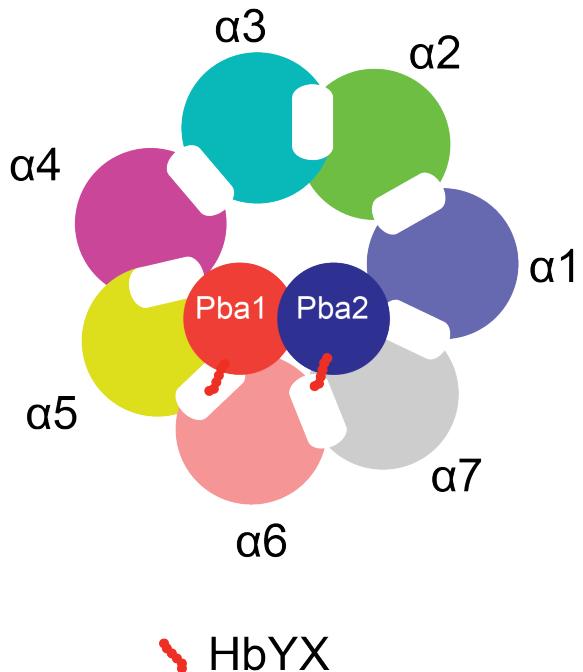
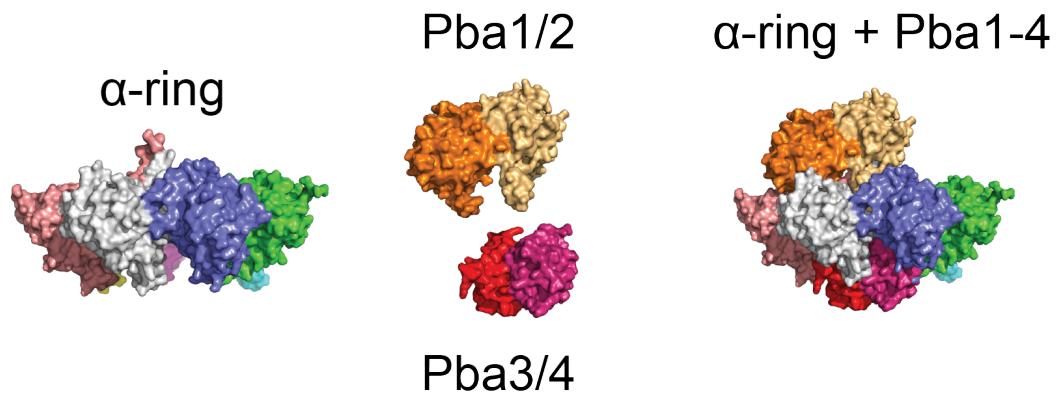


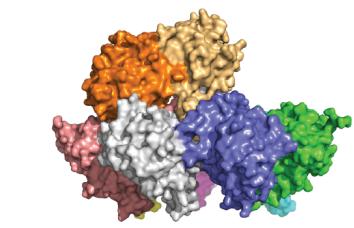
Figure 1.8: Pba1 and Pba2 binding sites via their HbYX motifs into adjacent α -subunit pairs. α -ring is shown with adjacent Pba1 and Pba2 binding surfaces that intercalate between α -subunits. Pba1 intercalates between α_5 and α_6 , while Pba2 intercalates between α_6 and α_7 (Yashiroda et al., 2008).

[Figure 1.9 caption follows on next page]

A

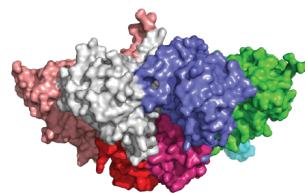


B



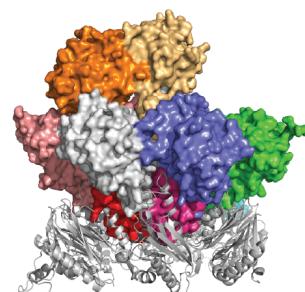
α -ring + Pba1/2

C



α -ring + Pba3/4

D



Pba3/4 occlude β 3 subunit binding

Figure 1.9 (preceding page): Structural Diagram of opposing Pba1/2 Pba3/4 binding. Structures were obtained from the Protein Data Bank with PDB ID 4G4S for Pba1 and Pba2 bound to yeast 20S proteasome (Stadtmueller et al., 2012), and PDB ID 2Z5C for Pba3 and Pba4 bound to the α 5 subunit(Yashiroda et al., 2008) and visualized using PyMOL (Schrodinger, 2015). The α 5 subunit was used to align the structures so that the opposing Pba1/2 and Pba3/4 interfaces could be visualized in reference to the 20S α -ring. **(A)** 20S α -ring with alone (left), Pba1/2 heterodimer, and Pba3/4 heterodimer alone (middle), and 20S α -ring with bound Pba1-4 shows the opposing binding surfaces, with Pba1/2 binding to the top of the α -ring, and Pba3/4 binding to the bottom of the α -ring where β -subunits normally bind in the full 20S proteasome. **(B)** α -ring with Pba1/2 binding. **(C)** α -ring with Pba3/4 binding. **(D)** Binding of Pba3/4 prevents premature binding of the β -ring (shown as a grey ribbon structure), specifically to occlude the β 3 subunit (Yashiroda et al., 2008).

1.5.2. 19S Regulatory Particle Assembly RP base

Like the chaperones that facilitate CP assembly, there are proteins that aide in proper formation of the RP. The majority of identified RP assembly chaperones are involved in the assembly of the RPT base ring. The RP base assembly chaperone Hsm3 was identified in the same screen for suppressors of Rad53 that identified the yeast CP assembly chaperones Pba1-4 (Le Tallec et al., 2007), although HSM3 was not characterized until two years later (Le Tallec et al., 2009). Several months later additional RP assembly chaperones (Nas2, Nas6 and Rpn14) were identified and characterized as components of an RP base assembly system (Funakoshi et al., 2009). The majority of these RP base chaperones function by binding the C-terminal domains of RPT subunits preventing them from forming pre-mature contacts with the CP (Funakoshi et al., 2009). However, their evolutionary history is likely distinct as they are made up of distinct domains including ARM/HEAT repeats in Hsm3, coiled-coil PDZ domains in Nas2, ankyrin repeats in Nas6, and WD40 repeats in Rpn14 (Funakoshi et al., 2009).

Immunoprecipitation of Hsm3 showed that it interacted with several RP base subunits; however no α -subunits or RP lid subunits were detected suggesting that Hsm3 may be specific to the RP base (Le Tallec et al., 2009). Fractionation of crude protein lysates from yeast showed that Hsm3 associated with free RP

but not with the mature 26S proteasome, suggesting that it may be involved in RP assembly (Le Tallec et al., 2009). Assembly intermediates containing Hsm3, Rpn1, and Rpt2 were identified when RP base assembly was impaired (Funakoshi et al., 2009). A mammalian ortholog of Hsm3, HSM3 was found to form a similar complex in mammals (Kaneko et al., 2009). Recently, a co-crystal structure involving Hsm3, Rpt1, and Rpt2 showed that Hsm3 forms contacts with the C-terminal domains of both Rpt1 and Rpt2, providing further support for the formation of this pre-assembly complex and that these assembly chaperones function to prevent premature binding of the C-terminus of RPT subunits to the CP (Barrault et al., 2012).

As described above, the C-terminal domains of RPT subunits contain HbYX motifs, which are important for binding the α -ring. Mutating the HbYX motif of either Rpt4 or Rpt6 by removing a single amino acid was sufficient to induce the formation of two separate pre-assembly complexes with one containing Hsm3 Rpn1, and Rpt1,2, and 5 and another containing Rpn14, Rpt4, and Rpt6 (Park et al., 2009). The formation of these pre-assembly complexes was blocked in triple null mutants for Hsm3, Bas6 and Rpn14 (Park et al., 2009). Interestingly single mutants in either Nas6, or Rpn14 did not exhibit strong assembly defects; however, yeast deficient in both Nas6, and Rpn14 exhibited a strong assembly defect suggesting

that Nas6, and Rpn14 play semi-redundant roles (Saeki et al., 2009). Despite this redundancy MS/MS analyses of tagged assembly chaperones and yeast-two-hybrid analyses showed Nas6 binding predominantly to Rpt3, while Rpn14 was found to predominantly bind to Rpt6 (Saeki et al., 2009). The yeast Nas2 assembly chaperone was found to bind to Rpt5, and its mammalian ortholog NAS2 was shown to form a complex with both RPT4 and RPT5. Recent structural analyses of Nas2 suggest that it binds the C-terminus of Rpt5 likely acting to prevent pre-mature binding with the CP (Satoh et al., 2014).

A general model emerges in which a majority of the RP base assembly chaperones act to prevent pre-mature association of unassembled RP base by actively binding the C-termini of RPT subunits (Tomko and Hochstrasser, 2013). This model involves four separate modules including [Hsm3 with Rpn1, Rpt1, and Rpt2], [Nas2 with Rpt4, and Rpt5], [Nas6 with Rpt3], and [Rpn14 with Rpt6]. While the exact ordered steps of RP base assembly are still unclear, the process appears to be conserved between both yeast and mammals. A model of RP base assembly using the yeast nomenclature is presented in 1.10 showing how the individual modules may come together to form the RP base.

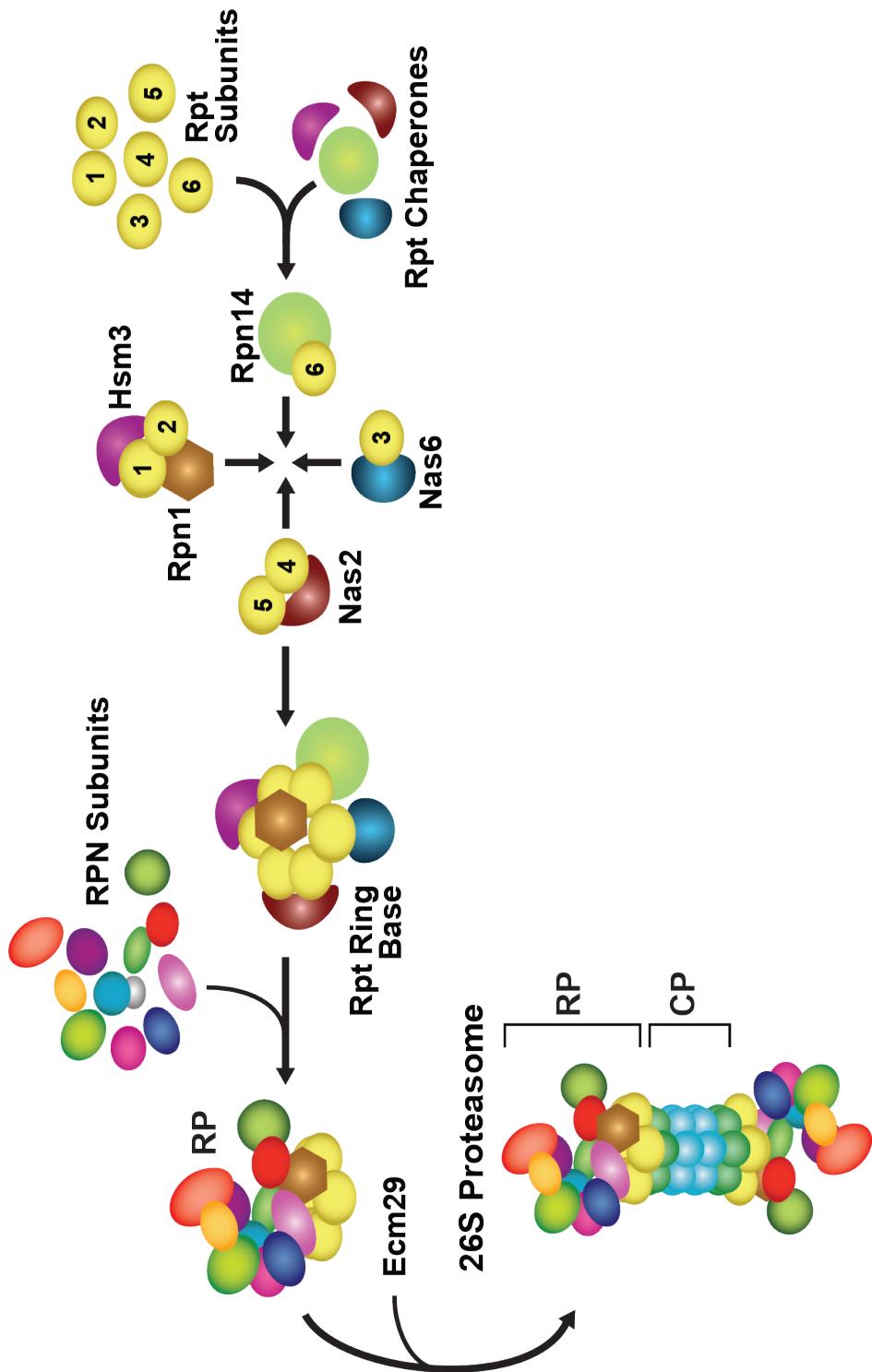


Figure 1.10: Schematic representation of RP Assembly. While the exact order of RP-base assembly is unclear, it involves several modules including [Hsm3 with Rpn1, Rpt1, and Rpt2], [Nas2 with Rpt4, and Rpt5], [Nas6 with Rpt3], and [Rpt14 with Rpt6].

1.5.3. 19S Regulatory Particle Assembly RP Lid

Several modules and chaperones are involved in the assembly of the RP base; however, construction of the RP lid is not well understood. RP lid assembly has primarily been investigated in yeast by mutant analyses of specific RP lid subunits including mutants in Rpn7, Rpn9 and Rpn12 (Fukunaga et al., 2010). MS/MS analyses of affinity purified lid complexes in these differing mutant backgrounds identified two different subassemblies, the first containing Rpn5,6,8,9, and 11, and the second containing Rpn3, Rpn7, and Sem1 (Fukunaga et al., 2010). Intriguingly, Sem1 was found to act as a bridge between Rpn3, and Rpn7, as cleavage of an engineered version of Sem1 containing a Tobacco Etch Virus (TEV)-cleavage site showed defects in RP lid assembly (Tomko and Hochstrasser, 2014). While Sem1 may play a role in proteasome biogenesis, alternative roles for Sem1 have been proposed. In the mature particle, Sem1 has been found to aide in recognition of ubiquitin chains which are linked at lysine 48 or 63 (Paraskevopoulos et al., 2014). Proteasomes affinity purified from Sem1 mutants had only mild defects in stability, suggesting that its main function may be that of ubiquitin recognition (Paraskevopoulos et al., 2014).

Structural studies of the RP lid, have identified conserved C-terminal helices that form a helical bundle important in assembly of this sub-complex (Estrin et al.,

2013). The assembly of the lid was investigated by expressing the full complement of yeast RP lid subunits in which the C-terminal helices were removed in varying combinations (Estrin et al., 2013). In the most extreme case, when the RPN6 helical bundle was omitted, a similar arrangement as described above was identified with four distinct sub-complexes identified that included [Rpn5, 8, 9, and 11], [Rpn3, and Rpn7], [Rpn6 alone], and [Rpn12 alone] (Estrin et al., 2013). Estrin *et al.* speculate on the exact ordering of the assembly, but this has not been thoroughly investigated.

While the exact ordering involving the other RP lid subunits is unclear, RPN12 is likely the last subunit to be added to the RP lid. In Rpn9 mutant cells, Rpn12 was completely absent from the RP lid (Fukunaga et al., 2010). However, mutants in Rpn12 exhibited the full complement of RP lid subunits suggesting that Rpn12 was likely the last subunit added to the RP lid (Tomko and Hochstrasser, 2011). The conserved C-terminus of RPN12 was found to contact both the RP base and lid suggesting that RPN12 might be a final checkpoint for RP lid and base assembly (Tomko and Hochstrasser, 2011). When the RP lid helical bundle described above is impaired, RPN12 cannot bind, suggesting a mechanism by which RPN12 can scan this bundle for proper assembly (Estrin et al., 2013).

1.5.4. RP-CP Assembly and Alternative Capping Particles

While there are dedicated chaperones for the CP and RP, an additional chaperone mediates assembly of the final particle (Ramos et al., 1998). ECM29 stabilizes the association of assembled CP and RP and provides a final quality control checkpoint for mature 26S proteasomes (Besche et al., 2014; Lehmann et al., 2010). Finally, in some situations, the RP is replaced entirely by alternate capping particles such as PA200 (also known as Blm10) or CDC48 (Barthelme and Sauer, 2012; Book et al., 2010; Schmidt et al., 2005). The functions of these caps are not yet clear, but recent proposals for PA200 have it participating in 26S proteasome assembly, helping shuttle proteasomes into the nucleus, and/or generating a ubiquitin-independent proteasome containing CP and PA200 only (Dange et al., 2011; Sadre-Bazzaz et al., 2010; Weberruss et al., 2013). However, in plants, mutants in PA200 exhibit no gross morphological defects showing that PA200 is not essential to plant growth and development (Book et al., 2010).

1.5.5. Proteasome Assembly in Plants

While much is known about proteasome assemble in both yeast and mammals, very little is known about how plants assemble this particle. There is some evidence that plants likely use a similar process as an ortholog of PAC2, PBAC2, has been

identified in affinity preparations for the plant CP complex (Book et al., 2010). A slightly different nomenclature is used in plants as compared to Pba in yeast and PAC in mammals as both of these can refer to proteasome α - and β -subunits (example, $\alpha 1$ is PAC1, and $\beta 1$ is PBA1 in *Arabidopsis*). Furthermore, PSI-BLAST-based bioinformatic analyses, similar to the one used to find orthologs between yeast and humans was able to identify some putative plant orthologs (PBAC1 and PBAC2) to the mammalian PAC1 and PAC2 assembly chaperones by two separate research groups, although their annotations in The *Arabidopsis* Information Resource (TAIR) are incomplete (Kusmierczyk et al., 2011; Le Tallec et al., 2007). These identified orthologs share very little sequence identity (<10%), although this is unsurprising given the poor sequence identity between the yeast and mammalian orthologs (Kusmierczyk et al., 2011; Le Tallec et al., 2007; Murata et al., 2009). Genomic searches based on BLAST alone, have failed to identify orthologs of the CP assembly chaperones; however putative orthologues for NAS2, NAS6, and HSM3 are readily identifiable in the genome (Chapter 3). Together these data suggested that our analysis of plant proteasome assembly has remained incomplete. One goal of this thesis, explored in Chapter 3, was to identify additional CP and RP interacting proteins, potentially including some of these assembly chaperones.

1.6. Proteasome Isotypes

In yeast, proteasome subunits are typically encoded as a single subunit, however in animals specific subunits are replaced to form different proteasome isotypes. This replacement typically occurs due to tissue specific expression of alternative CP subunits. Several examples of isotypes exist, including the Immunoproteasome, the thymoproteasome proteasome, and the testes-specific proteasome.

1.6.1. Immunoproteasome

While the proteasome was first discovered in the 1970's it wasn't until 20 years later that an alternative form of the proteasome, the immunoproteasome was discovered (Brown et al., 1991; Glynne et al., 1991; Ortiz-Navarrete et al., 1991). The immunoproteasome was discovered while trying to determine how small antigen presenting peptides are produced for the major histocompatibility complex (MHC). The MHC is a set of cell surface proteins involved in the acquired immune response (reviewed in (Neefjes et al., 2011)). Small peptides resulting from normally present cytosolic proteins (MHC class I antigens) are loaded into the MHC and presented on a cell's surface helping establish a normal profile of surface peptides recognized by the immune system. Atypical protein production, whether through viral or bacterial infection, or by cells that are proliferating abnormally, such as cancerous

cells, result in atypical surface peptides causing the immune system to respond and recruit CD8+ T-killer cells to remove the aberrant cell. Several of the identified genes involved in this process were termed low molecular weight proteins, or LMPs and their protein sequences were found to be related to the proteasome. The cytokine Interferon-gamma, a small signaling protein involved in the immune response was found to induce expression of three alternative β -subunits of the proteasome (β 1i, β 2i, and β 5i), which are then incorporated to form the immunoproteasome (1.11) (Aki et al., 1994; Groettrup et al., 1996).

An elegant study that compared the enzymatic activity of affinity purified 20S proteasomes with immunoproteasomes using model peptide substrates determined that cleavage after acidic residues via the β 1s PGPH activity is nearly eradicated by the replacement of the β 1 subunit with the immunoproteasome β 1is subunit which favors chymotryptic activities (Dahlmann et al., 2000). Together these altered enzymatic activities allow for the production of peptides more favorable for antigen presentation to the MHC.

1.6.2. Thymoproteasome proteasome

A sub-type of the immunoproteasome was discovered in the thymus. A search for proteasome related genes identified an uncharacterized ortholog of the β -subunit

that was preferentially expressed in thymus tissue, termed β 5t (Murata et al., 2007). This thymoproteasome was similar in composition to the immunoproteasome in that it contained the altered β 1i, and β 2i subunits; however, the β 5i subunit was replaced with this novel β 5t subunit (shown in 1.11) (Murata et al., 2007). Expression of the β 5t was even further sub-localized to cells in the thymic cortex, and mice lacking this subunit could not generate CD8+ T cells, normally involved in aiding in the destruction of foreign invaders (Murata et al., 2007).

1.6.3. Testes-specific proteasome

Like the discovery of the thymoproteasome proteasome, the discovery of the testes-specific isotype was initially driven by a bioinformatics search for proteasome related genes (Uechi et al., 2014). A search revealed an uncharacterized ortholog of the α 4 subunit that shared 85% amino acid identity, termed here α 4s (Uechi et al., 2014). Analysis of α 4ss expression by qRT-PCR analysis showed that it was expressed in all tissues; however only α 4s protein accumulated in testes (Uechi et al., 2014). α 4s is preferentially incorporated over the α 4 subunit in testes, however α 4s specific proteasomes did not show altered enzymatic activities, and its biological function remains unknown (Uechi et al., 2014).

1.6.4. α 4- α 4 proteasome Isotype

When defects are present in the Pba(PAC)3/4 heterodimer the α 3 subunit is replaced by the α 4 subunit, making an α 4- α 4 specific proteasome isotype (Kusmierczyk et al., 2008; Padmanabhan et al., 2016). Intriguingly, this α 4- α 4 isotype occurs in both mammals and yeast, conferring increased resistance to heavy metal ions, suggesting an evolutionarily conserved role for these proteasome isotypes (Kusmierczyk et al., 2008; Padmanabhan et al., 2016). Additionally, α 4 levels are modulated by several oncogenes including BRCA1 (Li et al., 2015), and the formation of these alternative subtypes is highly dependent on the relative availability of α 3 vs α 4. Together these data suggest that α 4- α 4 proteasome isotypes may play a role in the progression of cancer (Padmanabhan et al., 2016).

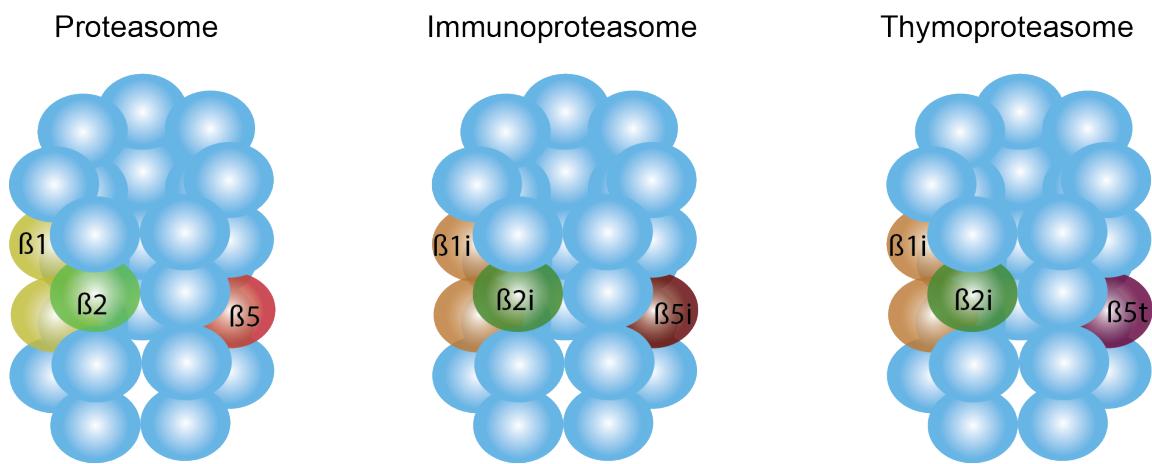


Figure 1.11: Representation of subunits replaced for both the immunoproteasome and the thyoproteasome. Adapted from (Groettrup et al., 2009). The three catalytically active sites of the proteasome $\beta 1$, $\beta 2$, and $\beta 5$ are replaced with the immunoproteasome variants $\beta 1i$, $\beta 2i$, and $\beta 5i$. Furthermore, in the thyoproteasome the $\beta 5i$ subunit is replaced with another variant the $\beta 5t$ subunit.

1.6.5. Plant Proteasome Isotypes

Outside of the proteasome isotypes discussed above, animal and yeast genomes typically encode a single gene for most proteasome subunits. In rice the RPT family has been duplicated such that four of the six RPT subunits are encoded as paralogs (Shibahara et al., 2004). While the function of these duplicated subunits is unclear, they do exhibit alternative tissue specific expression patterns as determined by MS/MS analyses using total protein extracts. Specifically, the expression patterns differed between bran and callus suggesting that there is tissue specific regulation of these sets of RPT subunits. In the most extreme case, most subunits in *Arabidopsis* are encoded as paralogs, likely due to a genome duplication (Book et al., 2010; Fu et al., 1998). MS/MS analyses of proteasomes purified from *Arabidopsis* showed that most of these subunit paralogs are incorporated into the proteasome. Therefore, *Arabidopsis* has the propensity to assemble its proteasomes into a diverse array of proteasome isotypes by incorporating specific subsets of isoforms. Some genetic evidence exists that suggests that these paralogs may have differing functions. RPN1a and RPN1b play differing roles in embryogenesis (Brukhin et al., 2005), and overexpression of RPN5a induces an early senescent phenotype while overexpression of RPN5b does not (Book et al., 2009). Despite these observations, it is unclear if plants assemble their proteasome subunit isoforms into compositionally and func-

tionally distinct isotypes. One of the objectives of this thesis was to determine if *Arabidopsis* can assemble proteasome isotypes by leveraging label-free quantitative mass spectrometric analysis of affinity preparations enriched for either the RPT4a or RPT4b subunit isoforms (explored in Chapter 3). If plants can assemble distinct isotypes, preparations for either isoform may show enrichment for other subunit isoforms; however, if plants assemble its subunit isoforms in a random fashion the preparations should be similar.

1.7. Conclusions

The *Arabidopsis* 26S proteasome exists *in planta* as a diverse array of complexes containing multiple subunit isoforms and interacting proteins (Book et al., 2010; Fu et al., 1999; Yang et al., 2004). However, there are several gaps in our knowledge of the plant particle. While the yeast and mammalian CP assembly systems are well described, very little is known about proteasome assembly in plants. Initial MS/MS analyses of the plant complex identified a likely ortholog related to PAC2, named PBAC2, suggesting that the plant system may be similar; however, no additional assembly chaperones were described (Book et al., 2010). Several putative *Arabidopsis* assembly orthologs were identified through PSI-BLAST analyses including orthologs for PAC1, and PAC2 by both the Hoschstrasser group, and others

(Kusmierczyk et al., 2011; Le Tallec et al., 2007); however, putative orthologs of PAC3, and PAC4 remained to be identified, until this thesis work. Taken together these data suggested that our analysis of the *Arabidopsis* CP interactome was incomplete and further analyses of CP associated proteins was necessary. Previous MS/MS analyses of the plant proteasome also failed to identify obvious orthologs of RP assembly chaperones. This suggested that our analyses of the RP interactome, like our previous analysis of the CP interactome, were also incomplete. One reason these previous analyses may have failed to identify RP assembly chaperones is because our affinity purifications targeted the CP specifically. To overcome this potential problem, I developed an affinity purification that targeted the RP through two separate isoforms of RPT4, RPT4a, and RPT4b. I exploited the fact that RPT4s N terminus was likely solvent accessible (Figure 1-12) and used a similar strategy developed previously by Book *et al.* (Book et al., 2010). Through quantitative MS/MS analyses I defined the plant RP interactome, as shown in Chapter 3. Additionally, I investigated the possibility of forming proteasome isotypes by the incorporation of specific subunit isoforms in combination with either RPT4a or RPT4b. In summary I identify CP- and RP-specific interactors, many of which may be putative orthologs of both the plant CP and RP assembly chaperones, and show that proteasomes enriched for either RPTa or RPT4b are largely similar suggesting that plants assemble

their proteasome subunit isoforms randomly.

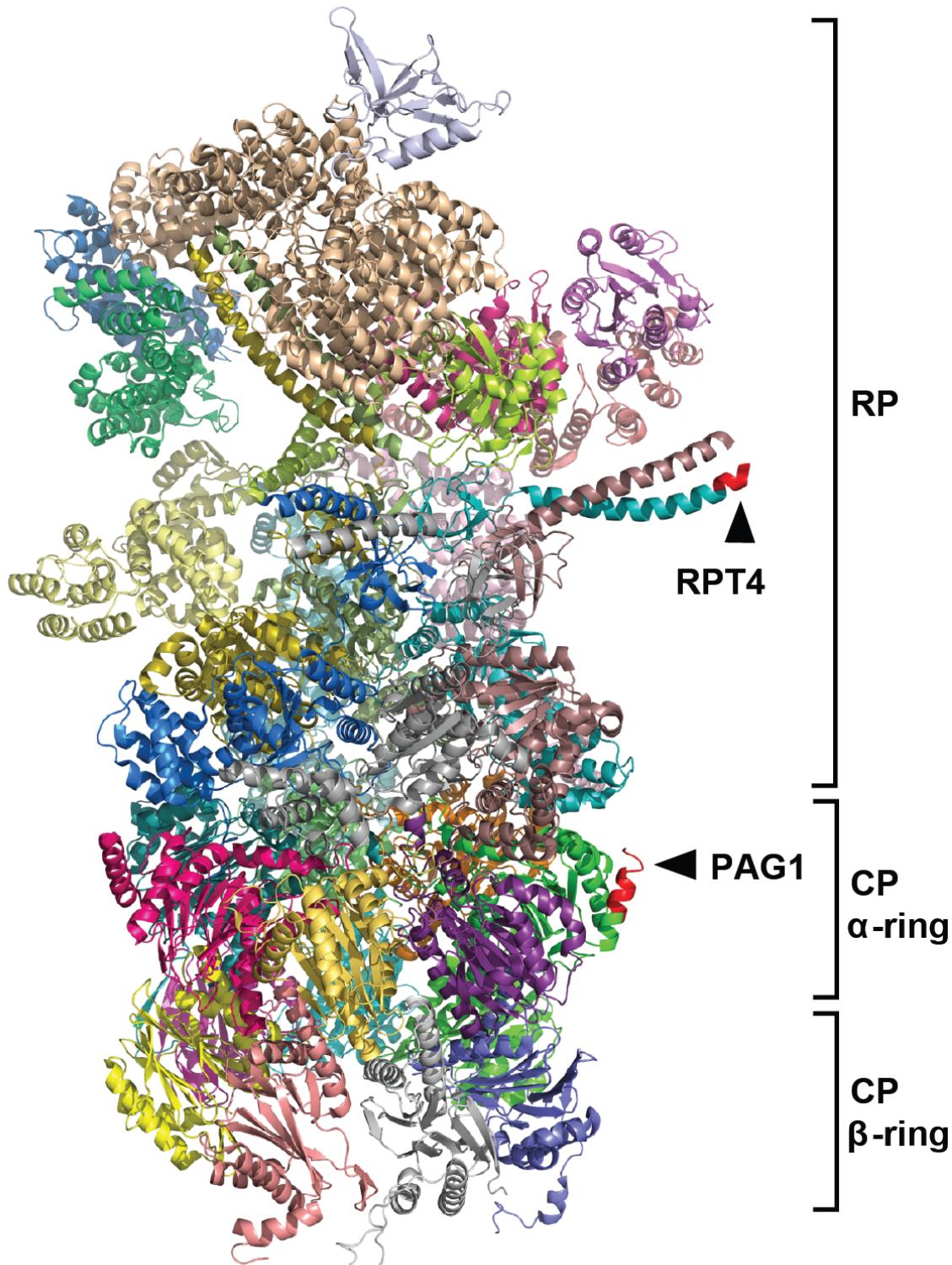


Figure 1.12: Solvent accessible tag locations on the 26S proteasome for both PAG1 and RPT4. A structural model of the 26S proteasome from yeast at subatomic resolution modified from PDB ID 4CR2 (Beck et al., 2012). The RP subunits, as well as the CP α and β -rings are shown. Highlighted in red, and indicated by black arrowheads, are the positions where FLAG affinity tags have been successfully used to enrich for *Arabidopsis* 26S proteasomes.

1.8. Literature Cited

- Aki, M., Shimbara, N., Takashina, M., Akiyama, K., Kagawa, S., Tamura, T., Tanahashi, N., Yoshimura, T., Tanaka, K., and Ichihara, A.** (1994). Interferon-gamma induces different subunit organizations and functional diversity of proteasomes. *J. Biochem.* **115**: 257–69.
- Angot, A., Peeters, N., Lechner, E., Vailleau, F., Baud, C., Gentzbittel, L., Sartorel, E., Genschik, P., Boucher, C., and Genin, S.** (2006). *Ralstonia solanacearum* requires F-box-like domain-containing type III effectors to promote disease on several host plants. *Proc. Natl. Acad. Sci. U.S.A.* **103**: 14620–5.
- Arendt, C.S.** (1999). Eukaryotic 20S proteasome catalytic subunit propeptides prevent active site inactivation by N-terminal acetylation and promote particle assembly. *The EMBO Journal* **18**: 3575–3585.
- Arendt, C.S. and Hochstrasser, M.** (1997). Identification of the yeast 20S proteasome catalytic centers and subunit interactions required for active site formation. *Proc. Natl. Acad. Sci. U.S.A.* **94**: 7156–7161.
- Armon, T., Ganoth, D., and Hershko, A.** (1990). Assembly of the 26S complex that degrades proteins ligated to ubiquitin is accompanied by the formation of ATPase activity. *J. Biol. Chem.* **265**: 20723–20726.
- Arrigo, A.P., Tanaka, K., Goldberg, A.L., and Welch, W.J.** (1988). Identity of the 19S prosome particle with the large multi-functional protease complex of mammalian cells (the proteasome). *Nature* **331**: 192–194.
- Barrault, M.B., Richet, N., Godard, C., Murciano, B., Le Tallec, B., Rousseau, E., Legrand, P., Charbonnier, J.B., Le Du, M.H., Guerois, R., Ochsenbein, F., and Peyroche, A.** (2012). Dual functions of the Hsm3 protein in chaperoning and scaffolding regulatory particle subunits during the proteasome assembly. *Proc. Natl. Acad. Sci. U.S.A.* **109**: E1001–10.
- Barthelme, D. and Sauer, R.T.** (2012). Identification of the Cdc48-20S proteasome as an ancient AAA proteolytic machine. *Science* **337**: 843–846.
- Baumeister, W., Dahlmann, B., Hegerl, R., Kopp, F., Kühn, L., and Pfeifer, G.** (1988). Electron microscopy and image analysis of the multi-catalytic proteinase. *FEBS Lett.* **241**: 239–245.

- Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J.M., Villa, E., Baumeister, W., and Förster, F.** (2012). Near-atomic resolution structural model of the yeast 26S proteasome. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 14870–14875.
- Berndsen, C.E. and Wolberger, C.** (2014). New insights into ubiquitin E3 ligase mechanisms. *Nat. Struct. Mol. Biol.* **21**: 301–307.
- Besche, H.C., Sha, Z., Kukushkin, N.V., Peth, A., Hock, E.M., Kim, W., Gygi, S.P., Gutierrez, J.A., Liao, H., Dick, L.R., and Goldberg, A.L.** (2014). Auto-ubiquitination of the 26S proteasome on Rpn13 regulates breakdown of ubiquitin conjugates. *EMBO J.* **33**: 1159–1176.
- Bhattacharyya, S., Yu, H., Mim, C., and Matouschek, A.** (2014). Regulated protein turnover: snapshots of the proteasome in action. *Nat. Rev. Mol. Cell Biol.* **15**: 122–133.
- Book, A.J., Gladman, N.P., Lee, S.S., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2010). Affinity purification of the *Arabidopsis* 26S proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* **285**: 25554–25569.
- Book, A.J., Smalle, J., Lee, K.H., Yang, P., Walker, J.M., Casper, S., Holmes, J.H., Russo, L.A., Buzzinotti, Z.W., Jenik, P.D., and Vierstra, R.D.** (2009). The RPN5 subunit of the 26S proteasome is essential for gametogenesis, sporophyte development, and complex assembly in *Arabidopsis*. *Plant Cell* **21**: 460–78.
- Brown, M.G., Driscoll, J., and Monaco, J.J.** (1991). Structural and serological similarity of MHC-linked LMP and proteasome (multicatalytic proteinase) complexes. *Nature* **353**: 355–7.
- Brukhin, V., Gheyselinck, J., Gagliardini, V., Genschik, P., and Grossniklaus, U.** (2005). The RPN1 subunit of the 26S proteasome in *Arabidopsis* is essential for embryogenesis. *Plant Cell* **17**: 2723–37.
- Ciechanover, A., Elias, S., Heller, H., Ferber, S., and Hershko, A.** (1980a). Characterization of the heat-stable polypeptide of the ATP-dependent proteolytic system from reticulocytes. *J. Biol. Chem.* **255**: 7525–8.
- Ciechanover, A., Finley, D., and Varshavsky, A.** (1984). Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. *Cell* **37**: 57–66.

- Ciechanover, A., Heller, H., Elias, S., Haas, A.L., and Hershko, A.** (1980b). ATP-dependent conjugation of reticulocyte proteins with the polypeptide required for protein degradation. *Proc. Natl. Acad. Sci. U.S.A.* **77**: 1365–1368.
- Clough, R.C. and Vierstra, R.D.** (1997). Phytochrome degradation. *Plant, Cell and Environment* **20**: 713–721.
- Dahlmann, B., Ruppert, T., Kuehn, L., Merforth, S., and Kloetzel, P.M.** (2000). Different proteasome subtypes in a single tissue exhibit different enzymatic properties. *J. Mol. Biol.* **303**: 643–53.
- Dange, T., Smith, D.M., Noy, T., Rommel, P.C., Jurzitz, L., Cordero, R.J., Legendre, A., Finley, D., Goldberg, A.L., and Schmidt, M.** (2011). The Blm10 protein promotes proteasomal substrate turnover by an active gating mechanism. *J. Biol. Chem.* **286**: 42830–42839.
- Dharmasiri, N., Dharmasiri, S., and Estelle, M.** (2005). The F-box protein TIR1 is an auxin receptor. *Nature* **435**: 441–445.
- Disch, S., Anastasiou, E., Sharma, V.K., Laux, T., Fletcher, J.C., and Lenhard, M.** (2006). The E3 ubiquitin ligase BIG BROTHER controls arabidopsis organ size in a dosage-dependent manner. *Curr. Biol.* **16**: 272–9.
- Elsasser, S., Gali, R.R., Schwickart, M., Larsen, C.N., Leggett, D.S., Müller, B., Feng, M.T., Tübing, F., Dittmar, G.A., and Finley, D.** (2002). Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat. Cell Biol.* **4**: 725–730.
- Estrin, E., Lopez-Blanco, J.R., Chacon, P., and Martin, A.** (2013). Formation of an intricate helical bundle dictates the assembly of the 26S proteasome lid. *Structure* **21**: 1624–35.
- Etlinger, J.D. and Goldberg, A.L.** (1977). A soluble, ATP-dependent proteolytic system responsible for degradation of abnormal proteins in reticulocytes. *Proc. Natl. Acad. Sci. U.S.A.* **74**: 54–58.
- Farmer, L.M., Book, A.J., Lee, K.H., Lin, Y.L., Fu, H., and Vierstra, R.D.** (2010). The RAD23 family provides an essential connection between the 26S proteasome and ubiquitylated proteins in Arabidopsis. *Plant Cell* **22**: 124–142.
- Fatimababy, A.S., Lin, Y.L., Usharani, R., Radjacommare, R., Wang, H.T., Tsai, H.L., Lee, Y., and Fu, H.** (2010). Cross-species divergence of the major recognition pathways of ubiquitylated substrates for ubiquitin/26S proteasome-mediated proteolysis. *FEBS J.* **277**: 796–816.

- Finley, D.** (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu. Rev. Biochem.* **78**: 477–513.
- Finley, D., Ciechanover, A., and Varshavsky, A.** (1984). Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. *Cell* **37**: 43–55.
- Finley, D., Ozkaynak, E., and Varshavsky, A.** (1987). The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. *Cell* **48**: 1035–1046.
- Frentzel, S., Pesold-Hurt, B., Seelig, A., and Kloetzel, P.M.** (1994). 20 S proteasomes are assembled via distinct precursor complexes. Processing of LMP2 and LMP7 proproteins takes place in 13-16 S preproteasome complexes. *J. Mol. Biol.* **236**: 975–81.
- Fu, H., Girod, P.A., Doelling, J.H., Van Nocker, S., Hochstrasser, M., Finley, D., and Vierstra, R.D.** (1999). Structural and functional analyses of the 26S proteasome subunits from plants. *Mol. Biol. Rep.* **26**: 137–146.
- Fu, H., Sadis, S., Rubin, D.M., Glickman, M.H., Van Nocker, S., Finley, D., and Vierstra, R.D.** (1998). Multi-ubiquitin chain binding and protein degradation are mediated by distinct domains within the 26S proteasome subunit MCB1. *J. Biol. Chem.* **273**: 1970–1981.
- Fujinami, K., Tanahashi, N., Tanaka, K., Ichihara, A., Cejka, Z., Baumeister, W., Miyawaki, M., Sato, T., and Nakagawa, H.** (1994). Purification and characterization of the 26S proteasome from spinach leaves. *J. Biol. Chem.* **269**: 25905–25910.
- Fukunaga, K., Kudo, T., Toh-e, A., Tanaka, K., and Saeki, Y.** (2010). Dissection of the assembly pathway of the proteasome lid in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **396**: 1048–53.
- Funakoshi, M., Tomko, R.J., Kobayashi, H., and Hochstrasser, M.** (2009). Multiple assembly chaperones govern biogenesis of the proteasome regulatory particle base. *Cell* **137**: 887–899.
- Ganoth, D., Leshinsky, E., Eytan, E., and Hershko, A.** (1988). A multi-component system that degrades proteins conjugated to ubiquitin: resolution of factors and evidence for ATP-dependent complex formation. *J. Biol. Chem.* **263**: 12412–12419.

- Glickman, M.H., Rubin, D.M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V.A., and Finley, D.** (1998a). A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3. *Cell* **94**: 615–623.
- Glickman, M.H., Rubin, D.M., Fried, V.A., and Finley, D.** (1998b). The regulatory particle of the *Saccharomyces cerevisiae* proteasome. *Mol. Cell. Biol.* **18**: 3149–3162.
- Glotzer, M., Murray, A.W., and Kirschner, M.W.** (1991). Cyclin is degraded by the ubiquitin pathway. *Nature* **349**: 132–138.
- Glynne, R., Powis, S.H., Beck, S., Kelly, A., Kerr, L.A., and Trowsdale, J.** (1991). A proteasome-related gene between the two ABC transporter loci in the class II region of the human MHC. *Nature* **353**: 357–60.
- Groettrup, M., Kirk, C.J., and Basler, M.** (2009). Proteasomes in immune cells: more than peptide producers? *Nat. Rev. Immunol.* **10**: 73–78.
- Groettrup, M., Kraft, R., Kostka, S., Standera, S., Stohwasser, R., and Kloetzel, P.M.** (1996). A third interferon-gamma-induced subunit exchange in the 20S proteasome. *Eur. J. Immunol.* **26**: 863–9.
- Groll, M., Bajorek, M., Köhler, A., Moroder, L., Rubin, D.M., Huber, R., Glickman, M.H., and Finley, D.** (2000). A gated channel into the proteasome core particle. *Nat. Struct. Biol.* **7**: 1062–1067.
- Groll, M., Heinemeyer, W., Jäger, S., Ullrich, T., Bochtler, M., Wolf, D.H., and Huber, R.** (1999). The catalytic sites of 20S proteasomes and their role in subunit maturation: a mutational and crystallographic study. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 10976–10983.
- Gu, Z.C. and Enenkel, C.** (2014). Proteasome assembly. *Cell. Mol. Life Sci.* **71**: 4729–4745.
- Hanna, J., Hathaway, N.A., Tone, Y., Crosas, B., Elsasser, S., Kirkpatrick, D.S., Leggett, D.S., Gygi, S.P., King, R.W., and Finley, D.** (2006). The deubiquitinating enzyme Ubp6 functions non-catalytically to delay proteasomal degradation. *Cell* **127**: 99–111.
- Hatfield, P.M. and Vierstra, R.D.** (1989). Ubiquitin-dependent proteolytic pathway in wheat germ: isolation of multiple forms of the E1 ubiquitin-activating enzyme. *Biochemistry (Mosc.)* **28**: 735–742.

- Heinemeyer, W., Fischer, M., Krimmer, T., Stachon, U., and Wolf, D.H.** (1997). The active sites of the eukaryotic 20S proteasome and their involvement in subunit precursor processing. *J. Biol. Chem.* **272**: 25200–25209.
- Hershko, A., Ciechanover, A., Heller, H., Haas, A.L., and Rose, I.A.** (1980). Proposed role of ATP in protein breakdown: conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc. Natl. Acad. Sci. U.S.A.* **77**: 1783–1786.
- Hirano, Y., Hayashi, H., Iemura, S.i., Hendil, K.B., Niwa, S.i., Kishimoto, T., Kasahara, M., Natsume, T., Tanaka, K., and Murata, S.** (2006). Cooperation of Multiple Chaperones Required for the Assembly of Mammalian 20S Proteasomes. *Mol. Cell* **24**: 977–984.
- Hirano, Y., Hendil, K.B., Yashiroda, H., Iemura, S.i., Nagane, R., Hioki, Y., Natsume, T., Tanaka, K., and Murata, S.** (2005). A heterodimeric complex that promotes the assembly of mammalian 20S proteasomes. *Nature* **437**: 1381–1385.
- Hirano, Y., Kaneko, T., Okamoto, K., Bai, M., Yashiroda, H., Furuyama, K., Kato, K., Tanaka, K., and Murata, S.** (2008). Dissecting beta-ring assembly pathway of the mammalian 20S proteasome. *EMBO J.* **27**: 2204–13.
- Hochstrasser, M., Ellison, M.J., Chau, V., and Varshavsky, A.** (1991). The short-lived MAT2 transcriptional regulator is ubiquitinated in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **88**: 4606–4610.
- Hough, R., Pratt, G., and Rechsteiner, M.** (1986). Ubiquitin-lysozyme conjugates: identification and characterization of an ATP-dependent protease from rabbit reticulocyte lysates. *J. Biol. Chem.* **261**: 2400–2408.
- Hough, R., Pratt, G., and Rechsteiner, M.** (1987). Purification of two high molecular weight proteases from rabbit reticulocyte lysate. *J. Biol. Chem.* **262**: 8303–8313.
- Hough, R. and Rechsteiner, M.** (1986). Ubiquitin-lysozyme conjugates: purification and susceptibility to proteolysis. *J. Biol. Chem.* **261**: 2391–2399.
- Hua, Z., Pool, J.E., Schmitz, R.J., Schultz, M.D., Shiu, S.H., Ecker, J.R., and Vierstra, R.D.** (2013). Epigenomic programming contributes to the genomic drift evolution of the F-Box protein superfamily in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **110**: 16927–16932.

- Hua, Z. and Vierstra, R.D.** (2011). The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* **62**: 299–334.
- Husnjak, K. and Dikic, I.** (2012). Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annu. Rev. Biochem.* **81**: 291–322.
- Husnjak, K., Elsasser, S., Zhang, N., Chen, X., Randles, L., Shi, Y., Hofmann, K., Walters, K.J., Finley, D., and Dikic, I.** (2008). Proteasome subunit Rpn13 is a novel ubiquitin receptor. *Nature* **453**: 481–488.
- Jager, S., Groll, M., Huber, R., Wolf, D.H., and Heinemeyer, W.** (1999). Proteasome beta-type subunits: unequal roles of propeptides in core particle maturation and a hierarchy of active site function. *J. Mol. Biol.* **291**: 997–1013.
- Kaneko, T., Hamazaki, J., Iemura, S., Sasaki, K., Furuyama, K., Natsume, T., Tanaka, K., and Murata, S.** (2009). Assembly pathway of the Mammalian proteasome base subcomplex is mediated by multiple specific chaperones. *Cell* **137**: 914–25.
- Katsir, L., Schilmiller, A.L., Staswick, P.E., He, S.Y., and Howe, G.A.** (2008). COI1 is a critical component of a receptor for jasmonate and the bacterial virulence factor coronatine. *Proc. Natl. Acad. Sci. U.S.A.* **105**: 7100–7105.
- Köhler, A., Cascio, P., Leggett, D.S., Woo, K.M., Goldberg, A.L., and Finley, D.** (2001). The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release. *Mol. Cell* **7**: 1143–1152.
- Kim, H.J., Oh, S.A., Brownfield, L., Hong, S.H., Ryu, H., Hwang, I., Twell, D., and Nam, H.G.** (2008). Control of plant germline proliferation by SCF(FBL17) degradation of cell cycle inhibitors. *Nature* **455**: 1134–7.
- Kisselev, A.F., van der Linden, W.A., and Overkleeft, H.S.** (2012). Proteasome inhibitors: an expanding army attacking a unique target. *Chem. Biol.* **19**: 99–115.
- Kremp, A., Schliephacke, M., Kull, U., and Schmid, H.P.** (1986). Prosomes exist in plant cells too. *Exp. Cell. Res.* **166**: 553–557.
- Kusmierczyk, A.R., Kunjappu, M.J., Funakoshi, M., and Hochstrasser, M.** (2008). A multimeric assembly factor controls the formation of alternative 20S proteasomes. *Nat. Struct. Mol. Biol.* **15**: 237–244.

- Kusmierczyk, A.R., Kunjappu, M.J., Kim, R.Y., and Hochstrasser, M.** (2011). A conserved 20S proteasome assembly factor requires a C-terminal HbYX motif for proteasomal precursor binding. *Nat. Struct. Mol. Biol.* **18**: 622–9.
- Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A.** (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**: 186–191.
- Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W.** (2012). Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 1380–1387.
- Le Tallec, B., Barrault, M.B., Courbeyrette, R., Guérois, R., Marsolier-Kergoat, M.C., and Peyroche, A.** (2007). 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and mammals. *Mol. Cell* **27**: 660–674.
- Le Tallec, B., Barrault, M.B., Guerois, R., Carre, T., and Peyroche, A.** (2009). Hsm3/S5b participates in the assembly pathway of the 19S regulatory particle of the proteasome. *Mol. Cell* **33**: 389–99.
- Lee, L.W., Moomaw, C.R., Orth, K., McGuire, M.J., DeMartino, G.N., and Slaughter, C.A.** (1990). Relationships among the subunits of the high molecular weight proteinase, macropain (proteasome). *Biochim. Biophys. Acta* **1037**: 178–85.
- Leggett, D.S., Glickman, M.H., and Finley, D.** (2005). Purification of proteasomes, proteasome sub-complexes, and proteasome-associated proteins from budding yeast. *Methods Mol. Biol.* **301**: 57–70.
- Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H.L., and Finley, D.** (2002). Multiple associated proteins regulate proteasome structure and function. *Mol. Cell* **10**: 495–507.
- Lehmann, A., Niewienda, A., Jechow, K., Janek, K., and Enenkel, C.** (2010). Ecm29 fulfills quality control functions in proteasome assembly. *Mol. Cell* **38**: 879–888.
- Li, D., Dong, Q., Tao, Q., Gu, J., Cui, Y., Jiang, X., Yuan, J., Li, W., Xu, R., Jin, Y., Li, P., Weaver, D.T., Ma, Q., Liu, X., and Cao, C.** (2015). c-Abl regulates proteasome abundance by controlling the ubiquitin-proteasomal degradation of PSMA7 subunit. *Cell Rep* **10**: 484–96.

- Li, X., Kusmierczyk, A.R., Wong, P., Emili, A., and Hochstrasser, M.** (2007). beta-Subunit appendages promote 20S proteasome assembly by overcoming an Ump1-dependent checkpoint. *EMBO J.* **26:** 2339–49.
- Lin, Y.L., Sung, S.C., Tsai, H.L., Yu, T.T., Radjacompare, R., Usharani, R., Fatimababy, A.S., Lin, H.Y., Wang, Y.Y., and Fu, H.** (2011). The defective proteasome, not substrate recognition function, is responsible for the null phenotypes of the *Arabidopsis* proteasome subunit RPN10. *Plant Cell* **23:** 2754–2773.
- Malik, M.N., Spivack, W.D., Sheikh, A.M., and Fenko, M.D.** (2004). The 26S proteasome in garlic (*Allium sativum*): purification and partial characterization. *J. Agric. Food. Chem.* **52:** 3350–3355.
- Marques, A.J., Glanemann, C., Ramos, P.C., and Dohmen, R.J.** (2007). The C-terminal extension of the beta7 subunit and activator complexes stabilize nascent 20 S proteasomes and promote their maturation. *J. Biol. Chem.* **282:** 34869–76.
- McClure, B.** (2004). S-RNase and SLF Determine S-Haplotype-Specific Pollen Recognition and Rejection. *The Plant Cell Online* **16:** 2840–2847.
- Más, P., Kim, W.Y., Somers, D.E., and Kay, S.A.** (2003). Targeted degradation of TOC1 by ZTL modulates circadian function in *Arabidopsis thaliana*. *Nature* **426:** 567–570.
- Murase, K., Hirano, Y., Sun, T.P., and Hakoshima, T.** (2008). Gibberellin-induced DELLA recognition by the gibberellin receptor GID1. *Nature* **456:** 459–63.
- Murata, S., Sasaki, K., Kishimoto, T., Niwa, S., Hayashi, H., Takahama, Y., and Tanaka, K.** (2007). Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science* **316:** 1349–53.
- Murata, S., Yashiroda, H., and Tanaka, K.** (2009). Molecular mechanisms of proteasome assembly. *Nat. Rev. Mol. Cell Biol.* **10:** 104–115.
- Murray, P.F., Giordano, C.V., Passeron, S., and Barneix, A.J.** (1997). Purification and characterization of the 20S proteasome from wheat leaves. *Plant Sci.* **125:** 127–136.
- Neefjes, J., Jongsma, M.L., Paul, P., and Bakke, O.** (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11:** 823–36.

- Nelson, C.J., Li, L., and Millar, A.H.** (2014). Quantitative analysis of protein turnover in plants. *Proteomics* **14**: 579–592.
- Ortiz-Navarrete, V., Seelig, A., Gernold, M., Frentzel, S., Kloetzel, P.M., and Hammerling, G.J.** (1991). Subunit of the '20S' proteasome (multicatalytic proteinase) encoded by the major histocompatibility complex. *Nature* **353**: 662–4.
- Ozaki, M., Fujinami, K., Tanaka, K., Amemiya, Y., Sato, T., Ogura, N., and Nakagawa, H.** (1992). Purification and initial characterization of the proteasome from the higher plant Spinacia oleracea. *J. Biol. Chem.* **267**: 21678–21684.
- Padmanabhan, A., Vuong, S.A., and Hochstrasser, M.** (2016). Assembly of an Evolutionarily Conserved Alternative Proteasome Isoform in Human Cells. *Cell Rep* **14**: 2962–74.
- Paraskevopoulos, K., Kriegenburg, F., Tatham, M.H., Rösner, H.I., Medina, B., Larsen, I.B., Brandstrup, R., Hardwick, K.G., Hay, R.T., Kragelund, B.B., Hartmann-Petersen, R., and Gordon, C.** (2014). Dss1 is a 26S proteasome ubiquitin receptor. *Mol. Cell* **56**: 453–461.
- Park, S., Roelofs, J., Kim, W., Robert, J., Schmidt, M., Gygi, S.P., and Finley, D.** (2009). Hexameric assembly of the proteasomal ATPases is templated through their C termini. *Nature* **459**: 866–70.
- Pathare, G.R., Nagy, I., Bohn, S., Unverdorben, P., Hubert, A., Körner, R., Nickell, S., Lasker, K., Sali, A., Tamura, T., Nishioka, T., Förster, F., Baumeister, W., and Bracher, A.** (2012). The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory sub-complexes together. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 149–154.
- Peters, J.M., Harris, J.R., and Kleinschmidt, J.A.** (1991). Ultrastructure of the approximately 26S complex containing the approximately 20S cylinder particle (multi-catalytic proteinase/proteasome). *Eur. J. Cell Biol.* **56**: 422–432.
- Rabl, J., Smith, D.M., Yu, Y., Chang, S.C., Goldberg, A.L., and Cheng, Y.** (2008). Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell* **30**: 360–368.
- Ramos, P.C., Höckendorff, J., Johnson, E.S., Varshavsky, A., and Dohmen, R.J.** (1998). Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. *Cell* **92**: 489–499.

- Ruschak, A.M., Religa, T.L., Breuer, S., Witt, S., and Kay, L.E.** (2010). The proteasome antechamber maintains substrates in an unfolded state. *Nature* **467**: 868–871.
- Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., and Coon, J.J.** (2013). Characterization and quantification of intact 26S proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One* **8**: e58157.
- Sadre-Bazzaz, K., Whitby, F.G., Robinson, H., Formosa, T., and Hill, C.P.** (2010). Structure of a Blm10 complex reveals common mechanisms for proteasome binding and gate opening. *Mol. Cell* **37**: 728–735.
- Saeki, Y., Toh, E.A., Kudo, T., Kawamura, H., and Tanaka, K.** (2009). Multiple proteasome-interacting proteins assist the assembly of the yeast 19S regulatory particle. *Cell* **137**: 900–13.
- Sakata, E., Bohn, S., Mihalache, O., Kiss, P., Beck, F., Nagy, I., Nickell, S., Tanaka, K., Saeki, Y., Förster, F., and Baumeister, W.** (2012). Localization of the proteasomal ubiquitin receptors Rpn10 and Rpn13 by electron cryo-microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **109**: 1479–1484.
- Sakata, E., Stengel, F., Fukunaga, K., Zhou, M., Saeki, Y., Förster, F., Baumeister, W., Tanaka, K., and Robinson, C.V.** (2011). The catalytic activity of Ubp6 enhances maturation of the proteasomal regulatory particle. *Mol. Cell* **42**: 637–649.
- Samach, A., Klenz, J.E., Kohalmi, S.E., Risseeuw, E., Haughn, G.W., and Crosby, W.L.** (1999). The UNUSUAL FLORAL ORGANS gene of *Arabidopsis thaliana* is an F-box protein required for normal patterning and growth in the floral meristem. *Plant J.* **20**: 433–45.
- Santner, A. and Estelle, M.** (2010). The ubiquitin-proteasome system regulates plant hormone signaling. *Plant J.* **61**: 1029–40.
- Satoh, T., Saeki, Y., Hiromoto, T., Wang, Y.H., Uekusa, Y., Yagi, H., Yoshihara, H., Yagi-Utsumi, M., Mizushima, T., Tanaka, K., and Kato, K.** (2014). Structural basis for proteasome formation controlled by an assembly chaperone nas2. *Structure* **22**: 731–43.
- Schellenberg, B., Ramel, C., and Dudler, R.** (2010). *Pseudomonas syringae* virulence factor syringolin A counteracts stomatal immunity by proteasome inhibition. *Mol. Plant. Microbe Interact.* **23**: 1287–93.

- Schliephacke, M., Kremp, A., Schmid, H.P., Köhler, K., and Kull, U.** (1991). Pro-somes (proteasomes) of higher plants. *Eur. J. Cell Biol.* **55**: 114–121.
- Schmid, H.P., Akhayat, O., Martins, C., Puvion, F., Köhler, K., and Scherrer, K.** (1984). The prosome: a ubiquitous morphologically distinct RNP particle associated with repressed mRNPs and containing specific scRNA and a characteristic set of proteins. *EMBO J.* **3**: 29–34.
- Schmidt, M., Haas, W., Crosas, B., Santamaria, P.G., Gygi, S.P., Walz, T., and Finley, D.** (2005). The HEAT repeat protein Blm10 regulates the yeast proteasome by capping the core particle. *Nat. Struct. Mol. Biol.* **12**: 294–303.
- Schreiner, P., Chen, X., Husnjak, K., Randles, L., Zhang, N., Elsasser, S., Finley, D., Dikic, I., Walters, K.J., and Groll, M.** (2008). Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453**: 548–552.
- Schrodinger, L.** (2015). The PyMOL Molecular Graphics System, Version 1.8.
- Shabek, N. and Zheng, N.** (2014). Plant ubiquitin ligases as signaling hubs. *Nat. Struct. Mol. Biol.* **21**: 293–296.
- Shanklin, J., Jabben, M., and Vierstra, R.D.** (1987). Red light-induced formation of ubiquitin-phytochrome conjugates: identification of possible intermediates of phytochrome degradation. *Proc. Natl. Acad. Sci. U.S.A.* **84**: 359–363.
- Shi, Y., Chen, X., Elsasser, S., Stocks, B.B., Tian, G., Lee, B.H., Shi, Y., Zhang, N., de Poot, S.A., Tuebing, F., Sun, S., Vannoy, J., Tarasov, S.G., Engen, J.R., Finley, D., and Walters, K.J.** (2016). Rpn1 provides adjacent receptor sites for substrate binding and deubiquitination by the proteasome. *Science* **351**.
- Shibahara, T., Kawasaki, H., and Hirano, H.** (2004). Mass spectrometric analysis of expression of ATPase subunits encoded by duplicated genes in the 19S regulatory particle of rice 26S proteasome. *Arch. Biochem. Biophys.* **421**: 34–41.
- Skoda, B. and Malek, L.** (1992). Dry pea seed proteasome: purification and enzymatic activities. *Plant Physiol.* **99**: 1515–1519.
- Smalle, J.A. and Vierstra, R.D.** (2004). The ubiquitin-26S proteasome proteolytic pathway. *Annu. Rev. Plant Biol.* **55**: 555–590.

- Smith, D.M., Kafri, G., Cheng, Y., Ng, D., Walz, T., and Goldberg, A.L.** (2005). ATP binding to PAN or the 26S ATPases causes association with the 20S proteasome, gate opening, and translocation of unfolded proteins. *Mol. Cell* **20**: 687–698.
- Sokolova, V., Li, F., Polovin, G., and Park, S.** (2015). Proteasome Activation is Mediated via a Functional Switch of the Rpt6 C-terminal Tail Following Chaperone-dependent Assembly. *Sci. Rep.* **5**: 14909.
- Stadtmueller, B.M., Kish-Trier, E., Ferrell, K., Petersen, C.N., Robinson, H., Myszka, D.G., Eckert, D.M., Formosa, T., and Hill, C.P.** (2012). Structure of a proteasome Pba1-Pba2 complex: implications for proteasome assembly, activation, and biological function. *J. Biol. Chem.* **287**: 37371–82.
- Stone, S.L., Anderson, E.M., Mullen, R.T., and Goring, D.R.** (2003). ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible Brassica pollen. *Plant Cell* **15**: 885–98.
- Tomko, R. J., J. and Hochstrasser, M.** (2011). Incorporation of the Rpn12 subunit couples completion of proteasome regulatory particle lid assembly to lid-base joining. *Mol. Cell* **44**: 907–17.
- Tomko, R. J., J. and Hochstrasser, M.** (2014). The intrinsically disordered Sem1 protein functions as a molecular tether during proteasome lid biogenesis. *Mol. Cell* **53**: 433–43.
- Tomko, R.J. and Hochstrasser, M.** (2013). Molecular architecture and assembly of the eukaryotic proteasome. *Annu. Rev. Biochem.* **82**: 415–445.
- Uechi, H., Hamazaki, J., and Murata, S.** (2014). Characterization of the testis-specific proteasome subunit alpha4s in mammals. *J. Biol. Chem.* **289**: 12365–74.
- Unverdorben, P., Beck, F., Ied, P., Schweitzer, A., Pfeifer, G., Plitzko, J.M., Baumeister, W., and Förster, F.** (2014). Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc. Natl. Acad. Sci. U.S.A.* **111**: 5544–5549.
- Van Nocker, S., Devereaux, Q., Rechsteiner, M., and Vierstra, R.D.** (1996). The Arabidopsis MBP1 gene encodes a conserved ubiquitin recognition component of the 26S proteasome. *Proc. Natl. Acad. Sci. U.S.A.* **93**: 856–860.

- Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., Koonin, E.V., and Deshaies, R.J.** (2002). Role of the Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**: 611–615.
- Vierstra, R.D.** (1993). Protein degradation in plants. *Annu. Rev. Plant Phys.* **44**: 385–410.
- Vierstra, R.D.** (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat. Rev. Mol. Cell Biol.* **10**: 385–397.
- Vierstra, R.D. and Sullivan, M.L.** (1988). Hemin inhibits ubiquitin-dependent proteolysis in both a higher plant and yeast. *Biochemistry (Mosc.)* **27**: 3290–3295.
- Wani, P.S., Suppahia, A., Capalla, X., Ondracek, A., and Roelofs, J.** (2016). Phosphorylation of the C-terminal tail of proteasome subunit alpha7 is required for binding of the proteasome quality control factor Ecm29. *Sci. Rep.* **6**: 27873.
- Waxman, L., Fagan, J.M., and Goldberg, A.L.** (1987). Demonstration of two distinct high molecular weight proteases in rabbit reticulocytes, one of which degrades ubiquitin conjugates. *J. Biol. Chem.* **262**: 2451–2457.
- Weber russ, M.H., Savulescu, A.F., Jando, J., Bissinger, T., Harel, A., Glickman, M.H., and Enenkel, C.** (2013). Blm10 facilitates nuclear import of proteasome core particles. *EMBO J.* **32**: 2697–2707.
- Wilkinson, K.D., Urban, M.K., and Haas, A.L.** (1980). Ubiquitin is the ATP-dependent proteolysis factor I of rabbit reticulocytes. *J. Biol. Chem.* **255**: 7529–7532.
- Witt, E., Zantopf, D., Schmidt, M., Kraft, R., Kloetzel, P.M., and Kruger, E.** (2000). Characterisation of the newly identified human Ump1 homologue POMP and analysis of LMP7(beta 5i) incorporation into 20 S proteasomes. *J. Mol. Biol.* **301**: 1–9.
- Woffenden, B.J., Freeman, T.B., and Beers, E.P.** (1998). Proteasome inhibitors prevent tracheary element differentiation in *Zinnia* mesophyll cell cultures. *Plant Physiol.* **118**: 419–429.
- Worden, E.J., Padovani, C., and Martin, A.** (2014). Structure of the Rpn11-Rpn8 dimer reveals mechanisms of substrate deubiquitination during proteasomal degradation. *Nat. Struct. Mol. Biol.* **21**: 220–227.

- Xie, Y. and Varshavsky, A.** (2000). Physical association of ubiquitin ligases and the 26S proteasome. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 2497–2502.
- Yanagawa, Y., Ohhashi, A., Murakami, Y., Saeki, Y., Yokosawa, H., Tanaka, K., Hashimoto, J., Sato, T., and Nakagawa, H.** (1999). Purification and characterization of the 26S proteasome from cultured rice (*Oryza sativa*) cells. *Plant Sci.* **149**: 33–41.
- Yang, P., Fu, H., Walker, J.M., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D.** (2004). Purification of the *Arabidopsis* 26S proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *J. Biol. Chem.* **279**: 6401–6413.
- Yashiroda, H., Mizushima, T., Okamoto, K., Kameyama, T., Hayashi, H., Kishimoto, T., Niwa, S., Kasahara, M., Kurimoto, E., Sakata, E., Takagi, K., Suzuki, A., Hirano, Y., Murata, S., Kato, K., Yamane, T., and Tanaka, K.** (2008). Crystal structure of a chaperone complex that contributes to the assembly of yeast 20S proteasomes. *Nat. Struct. Mol. Biol.* **15**: 228–36.
- Yoshimura, T., Kameyama, K., Takagi, T., Ikai, A., Tokunaga, F., Koide, T., Tanahashi, N., Tamura, T., Cejka, Z., Baumeister, W., Tanaka, K., and Ichihara, A.** (1993). Molecular characterization of the 26S proteasome complex from rat liver. *J. Struct. Biol.* **111**: 200–211.
- Zhang, Y., Zhao, Z., and Xue, Y.** (2009). Roles of Proteolysis in Plant Self-Incompatibility. *Annu. Rev. Plant Biol.* **60**: 21–42.
- Zhao, L., Huang, J., Zhao, Z., Li, Q., Sims, T.L., and Xue, Y.** (2010). The Skp1-like protein SSK1 is required for cross-pollen compatibility in S-RNase-based self-incompatibility. *Plant J.* **62**: 52–63.

Chapter 2

MORPHEUS SPECTRAL COUNTER: A COMPUTATIONAL TOOL

FOR LABEL-FREE QUANTITATIVE MASS SPECTROMETRY

USING THE MORPHEUS SEARCH ENGINE

2.1. Summary

Label-free quantitative MS based on the Normalized Spectral Abundance Factor (NSAF) has emerged as a straightforward and robust method to determine the relative abundance of individual proteins within complex mixtures. Here, we present Morpheus Spectral Counter (MSpC) as the first computational tool that directly calculates NSAF values from output obtained from Morpheus, a fast, open-source, peptide-MS/MS matching engine compatible with high-resolution accurate-mass instruments. NSAF has distinct advantages over other MS-based quantification methods, including a higher dynamic range as compared to isobaric tags, no requirement to align and re-extract MS1 peaks, and increased speed. MSpC features an easy to use graphic user interface that additionally calculates both distributed and unique NSAF values to permit analyses of both protein families and isoforms/proteoforms. MSpC determinations of protein concentration were linear over

several orders of magnitude based on the analysis of several high-mass accuracy datasets either obtained from PRIDE or generated with total cell extracts spiked with purified *Arabidopsis* 20S proteasomes. The MSpC software was developed in C# and is open sourced under a permissive license with the code made available at http://dcgemperline.github.io/Morpheus_SpC/.

2.2. Main Text

Quantification of individual polypeptides within complex mixtures by MS is an extremely useful tool to understand proteomic changes in organisms during growth and development, and after environmental perturbation (Wong and Cagney, 2010). While a number of MS/MS strategies have been developed to measure protein abundance, including Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC), labeling with isobaric tags, and Absolute Quantification of proteins (AQUA) (Gerber et al., 2003; Ong et al., 2002; Ross et al., 2004; Thompson et al., 2003), label-free quantification (LFQ) have become increasingly popular given their simplicity and low cost (Wong and Cagney, 2010; Zhang et al., 2006). One LFQ strategy infers abundance from the number of observed peptide spectra matches (PSMs). For these PSM-based approaches, changes in protein abundance can be generated artificially when total PSMs differ among samples and because longer proteins

tend to produce more raw counts. For these reasons normalizing for both protein length and total PSMs is paramount. While this adjustment can be made in a number of ways; one of the most straight forward methods is to use Normalized Spectral Abundance Factor (NSAF), a length- and count-normalized measure for each protein (Zybailov et al., 2006). Further improvements to the NSAF algorithm have been made by accounting for shared peptides in distributed NSAF (dNSAF), which distributes common PSMs among a family of isoforms/proteoforms based on the number of distinct PSMs observed for each isoform/proteoform, and unique NSAF (uNSAF), which ignores shared PSMs and only assigns distinct PSMs to each specific isoform/proteoform (Zhang et al., 2010).

The Morpheus MS search engine was recently designed for high-resolution, accurate-mass data obtained from Orbitrap-based instruments to provide faster matching of spectra to peptides (Wenger and Coon, 2013). Unfortunately, no downstream automated tools are available to facilitate LFQ analysis, which can be quite challenging, if not impossible, to complete manually when accounting for shared peptides. To overcome this bottleneck, we developed Morpheus Spectral Counter (MSpC) as the first LFQ computational tool that integrates directly with Morpheus to calculate NSAF, dNSAF, uNSAF, and corrected PSM (Fermin et al., 2011) values in complex protein samples. MSpC is fully automated, and only requires a

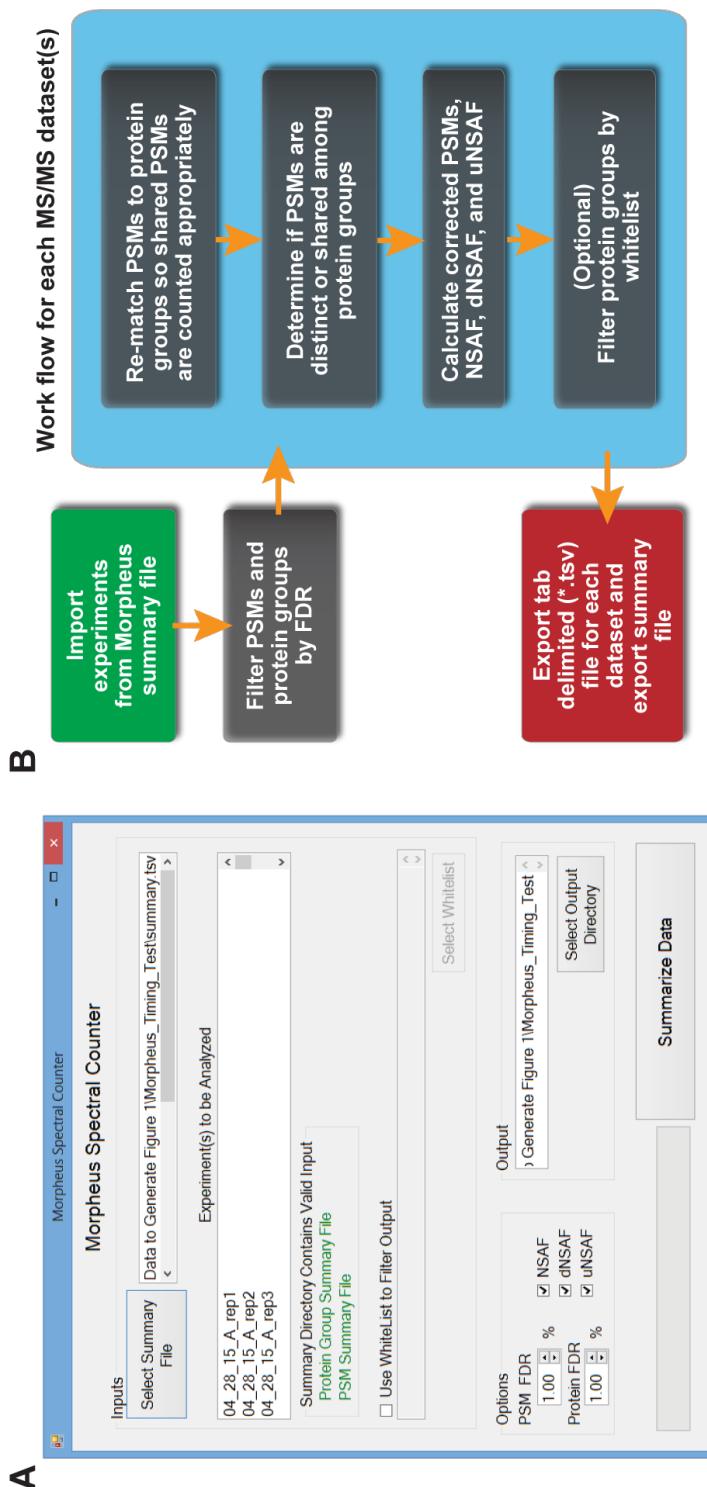


Figure 2.1: MSpC Graphic User Interface (GUI) and software flow chart. (A) Screenshot of the GUI. The input requires the user to select a Morpheus search summary file containing experiments to be analyzed. The user can optionally select a whitelist to filter output, and select an output directory. Additional options can set peptide and protein FDR cutoffs, and method of quantification for output, including Normalized Spectral Abundance Factor (NSAF), distributed NSAF (dNSAF), and unique NSAF (uNSAF). A progress bar highlights completion of the analysis. (B) Data analysis flow chart. Experiments and groups of experiments to be analyzed are imported through the Morpheus summary file. PSMs and protein groups are filtered at the specified FDR cutoff with a default of 1%.

Morpheus search summary file (summary.tsv) as input. The user interface (see Figure 2.1 (A)) allows one to select the summary file and displays the raw MS/MS files that will be analyzed by MSpC. Due to shared peptides being attributed to only one instance of a protein group in Morpheus's PSM file, PSMs are re-matched to all possible protein groups. PSMs are then cataloged as shared or as unique (distinctly matching one protein group) to generate NSAF, dNSAF, and uNSAF outputs. Finally, the output can be filtered for proteins of interest by specifying a comma delimited file containing unique identifiers and descriptions. Some important features of MSpC are its ability to handle fractionation experiments as input, and the ability to whitelist proteins of interest in the output by specifying a csv file (see Tutorial 2.5). Options exist to specify global PSM and protein group FDR rates (thus avoiding increased FDRs when one analyzes many experiments at once), to output NSAF, dNSAF, and uNSAF values, to require a minimum number of unique peptides to quantify a protein, and to specify an output directory. A progress bar indicates completion of the analysis by MSpC. To validate the accuracy of MSpC, we analyzed two MS/MS datasets available in PRIDE that were previously generated by high-energy collision-induced dissociation using Thermo Q-Exactive Orbitrap instruments.

Here, *Xenopus* egg (see top, Figure 2.2) and embryo (bottom, Figure 2.2) extracts

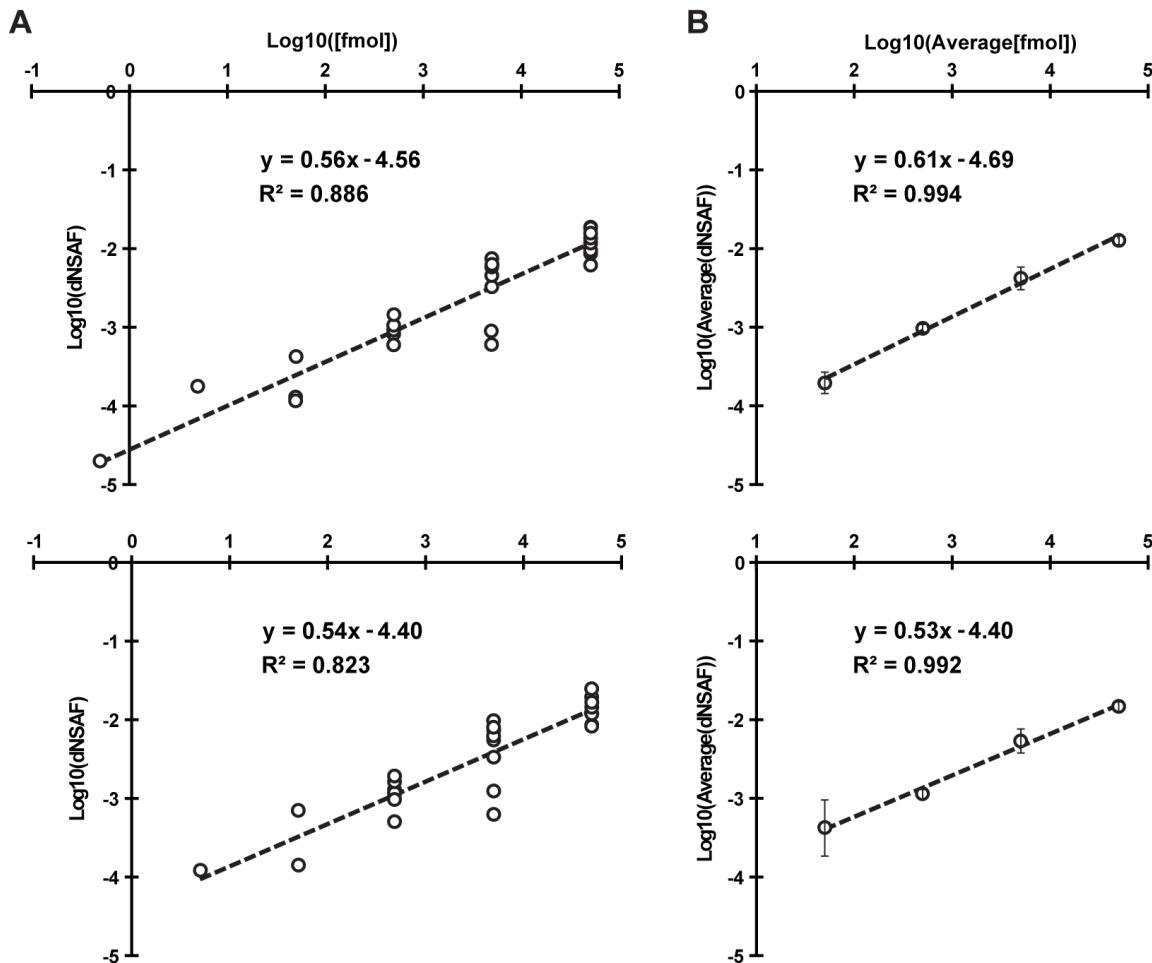


Figure 2.2: Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2). The array of UPS2 standards were spiked into *Xenopus laevis* egg (**Top**) and embryo (**Bottom**) extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein were determined by Morpheus and MSpC. (**A**) A log-log plot of dNSAF versus concentration for each UPS2 protein detected across each fmol range. (**B**) A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

were spiked at a 4:1 ratio with the Universal Proteome Standard 2 (UPS2), a mix of 48 purified proteins at defined molar ratios of 0.5, 5, 50, 500, 5000, and 50,000, with each ratio containing a different set of 8 of the 48 proteins. As shown in Figure 2.2 A, when the Morpheus/MSpC pipeline was used to calculate the average dNSAF value for each UPS2 protein, requiring only a single unique peptide to quantify, strong linear correlations ($R^2 = 0.886$ and 0.823) were obtained across a 1,000 fold change in abundance (50 fmol to 50,000 fmol). In fact, the R^2 values were similar to those obtained by others with PSM-based LFQ methods (Cox et al., 2014; Tu et al., 2014). This linear correlation was further strengthened when the dNSAF values were averaged for all UPS2 proteins within each of the concentration groups, with R^2 values of 0.994 and 0.992 for the egg and embryo datasets, respectively (see Figure 2.2 B). Notably, the slope of the concentration series was significantly less than unity, showing that NSAF measurements are not appropriate for absolute quantification, which was expected given that NSAF is a relative value.

We also reprocessed the UPS2 dataset using the option of requiring a minimum of two unique peptides for quantification, which should improve stringency. This option provided only a minor improvement in overall linearity for the average UPS2 dNSAF values, but decreased linearity when each UPS2 protein was considered individually and removed some UPS2 proteins at low concentrations (compare

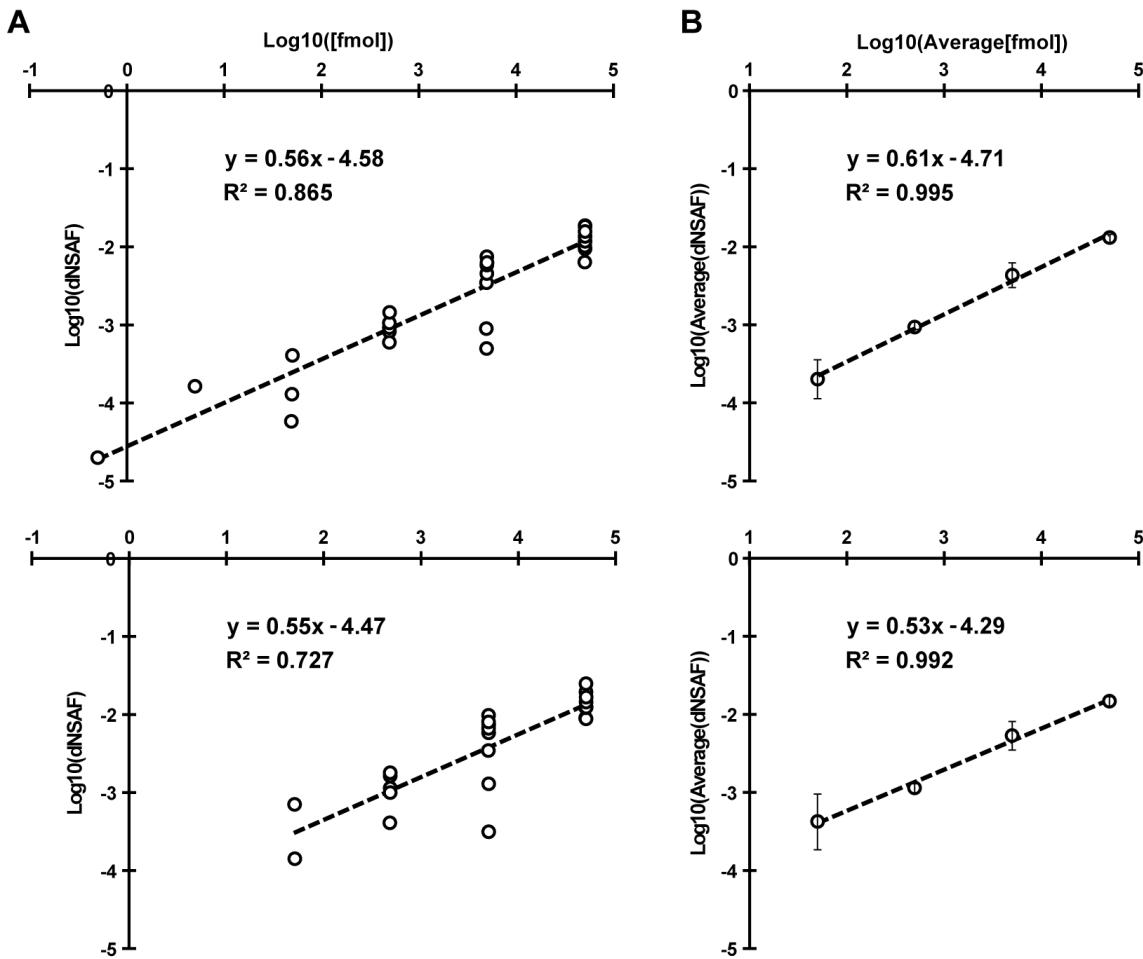
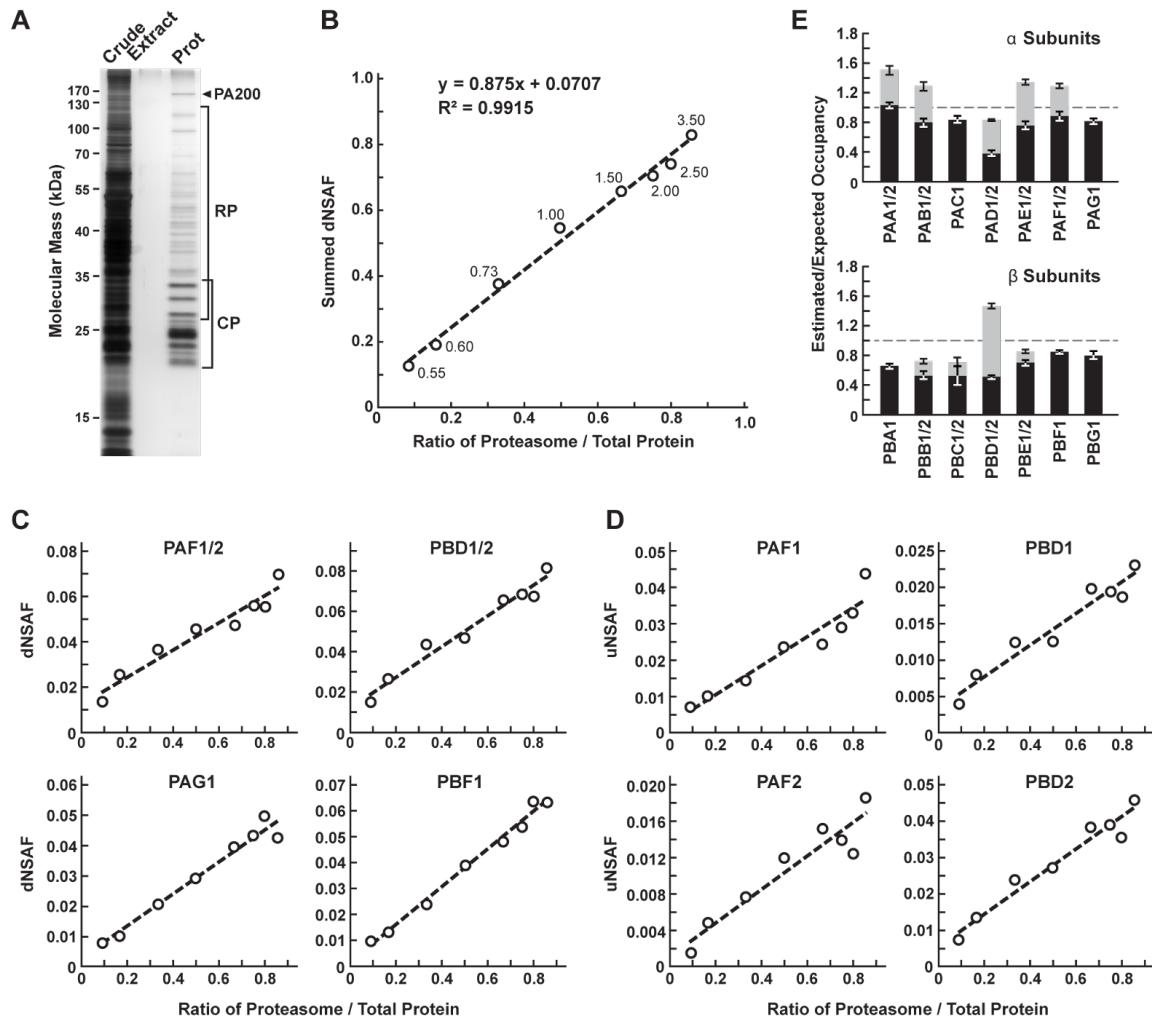


Figure 2.3: Re-Analysis of MS/MS datasets generated with the Universal Proteome Standard 2 (UPS2). The array of UPS2 standards were spiked into *Xenopus laevis* egg (**Top**) and embryo (**Bottom**) extracts at a range of concentrations. Following MS/MS analysis, dNSAF values for each protein were determined by Morpheus and MSpC with a change from Figure 2.2 in that two unique peptides were required to quantify a protein. **(A)** A log-log plot of dNSAF versus concentration for each UPS2 protein detected across each fmol range. **(B)** A log-log plot of average dNSAF vs average concentration of each group of UPS2 proteins at each fmol range: (50, 500, 5000, and 50,000 fmol).

Figure 2.3 A to Figure 2.2 A). Consequently, caution should be exercised when selecting this option even though it might provide a slight improvement in stringency (see Discussion on Requiring Two Peptide 2.4). To demonstrate the utility and accuracy of MSpC as applied to our work, we analyzed 20S proteasomes isolated from *Arabidopsis thaliana*. This particle contains multiple subunits assembled in stoichiometric amounts, with many subunits encoded by two paralogous genes of sufficient amino acid identity (typically >90% (Yang et al., 2004)) such that discrimination between paralogs can be challenging using LFQ approaches (Book et al., 2010). To simulate changes in 20S proteasome abundance, we added varying amounts of trypsinized proteasomes (0.05 µg to 3 µg) to a fixed amount of trypsinized *Escherichia coli* lysate (0.5 µg) to generate proteasome/lysate ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, 0.857. The digests were then subjected to MS/MS and the dNSAF value for each subunit along with the uNSAF value for individual isoforms were calculated by the Morpheus/MSpC pipeline (see Methods 2.3). The data from this experiment are deposited in PRIDE with ID PXD003002. As shown in Figure 2.4, MSpC provided an excellent determination for the overall abundance of 20S proteasomes within a complex mixture, along with a good reflection of the abundance of individual subunits and their isoforms.

When the dNSAF values for all subunits for the *Arabidopsis* 20S proteasome

[Figure 2.4 caption follows on next page]



including their isoforms (representing 14 distinct subunits, 10 of which exist as isoform pairs) were summed, a very close approximation of the dNSAF/actual abundance was obtained (slope=0.875) with a very strong linear correlation ($R^2 = 0.99$) over a ~10-fold range in protein abundance. When each 20S proteasome subunit was analyzed individually, a strong linear response was also obtained ($R^2 > 0.90$) for a majority of subunits (Figure 2.4 C and Table 2.1).

For example, reasonably accurate concentration plots were obtained for the

Figure 2.4 (preceding page): Confirmation of MSpC accuracy by analysis of MS/MS datasets generated with affinity purified *Arabidopsis* 20S proteasomes spiked into a total cell lysate from *E. coli*. Following MS/MS analysis, the dNSAF and uNSAF values for each subunit/isoform were determined by Morpheus and MSpC. (A) A silver-stained SDS-PAGE gel of 20S proteasome samples affinity purified from 10-d-old *Arabidopsis* seedlings. The crude seedling extract(CE), sample buffer (SB), and affinity-purified 20S proteasome samples (Prot) are shown. (B) Quantification of trypsinized 20S proteasomes when mixed at varying ratios with trypsinized total protein lysates from *E. coli*. The spiked samples were subjected to MS/MS followed by data analysis with the Morpheus and MSpC. dNSAF values for each proteasome subunit were averaged across three technical replicates, then summed to obtain an estimate of abundance for the 20S proteasome, and plotted against their known ratios. The total protein load is listed at each point in ug. (C and D) dNSAF and uNSAF values determined from the data in panel B for individual subunits (C) and their isoforms (D) for several subunits of the 20S proteasome. (E) Quantification accuracy of the Morpheus/MSpC pipeline for determining the amount of each α and β subunit of the 20S proteasome. Single subunit isoforms are in black, whereas subunits having two isoforms are shown in black and grey to reflect the contributions of isoforms 1 and 2 respectively. Each bar represents the average of eight technical replicates (\pm SE). The dashed line represents the expected value of one assuming an equal stoichiometry of each subunit within the particle.

PAF (α 6) and PBD (β 4) subunits that are encoded by the PAF1/2 and PBF1/2 gene pairs, and for the PAG (α 7) and PBF (β 6) subunits that are encoded by single PAG1 and PBF1 genes (R^2 from 0.94 to 0.99). Even when we calculated uNSAF values for individual isoforms added to the *E. coli* lysate, strong linear responses were obtained (e.g., the PAF1/PAF2 and PBD1/PBD2 pairs) with robust correlations (R^2 from 0.89 to 0.95) (Figure fig:proteasomespike D). Taken together, MSpC worked well for relative LFQ analysis of a multi-subunit complex and its individual subunits and isoforms within a complex proteomic mixture.

The Morpheus/MSpC pipeline also allowed us to calculate the respective incorporation of each paralog in the complex (see Isoform Incorporation Methods 2.3.4). As shown in Figure 2.4E, these estimated/expected occupancies were close to unity for most subunits within both the α and β rings of the 20S proteasome. The only strong deviation was for PBD1/2 (β 4), which had a greater dNSAF value relative to other β subunits across the experiments analyzed (see Table 2.1). The calculations for uNSAF values also estimated the relative proportion of each isoform within the complex for those subunits expressed from paralogous genes. The data obtained are similar to prior studies of the complex involving quantitative top-down proteomic analysis of purified proteasome samples using ultra violet-intrinsic fluorescence to quantify tyrosine-containing subunits (Russell et al., 2013). However, our MSpC

analysis provided a more complete picture as several subunit isoforms were difficult to quantify by fluorescence either because they lacked tryosine, or because their fluorescence peaks overlapped with those of other subunits/isoforms. Notably , the protein isoform ratios measured here agree well with the expression ratios for the paralogous genes (Book et al., 2010), suggesting that the protein isoform abundance generally reflects the relative transcriptional activity of the gene pair. We consistently estimated slightly more α ring subunits (PAA-PAG) versus β ring subunits (PBA-PBG) in the final MSpC calculations (Figure 2.4 E). This deviation could represent enhanced detection of α ring versus β ring subunits, or more likely that purification via the tagged α ring subunit PAG1 also isolated assembly intermediates comprised of only α ring subunits.

We compared the Morpheus and MSpC pipeline to the next most comparable open source, spectral-count-based LFQ pipeline, The Trans Proteomic Pipeline (TPP) (Deutsch et al., 2010) and ABACUS (Fermin et al., 2011) using our datasets generated with the 20S proteasome/*E. coli* lysate mixture (see Tables 2.1 and 2.2). Morpheus/MSpC slightly outperformed TPP/ABACUS by having a greater overall accuracy (average linearity of 0.88 compared to 0.84), and by having more subunits showing an R^2 linear correlation greater than 0.9 (14/23 subunits for MSpC versus and 11/23 for ABACUS).

Table 2.1: Table of dNSAF values for each 20S proteasome subunit generated by analyzing the proteasome spike in experiments with the Morpheus and MSpC pipeline. The top half of the table lists α 1-7 (PAA-PAG) where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists β 1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms)

Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PAA1	0.0102	0.0180	0.0275	0.0392	0.0505	0.0489	0.0467	0.0530		0.952
PAA2	0.0063	0.0072	0.0118	0.0121	0.0139	0.0218	0.0351	0.0195		0.652
PAB1	0.0091	0.0117	0.0209	0.0391	0.0377	0.0377	0.0268	0.0397		0.721
PAB2	0.0017	0.0062	0.0147	0.0223	0.0295	0.0277	0.0303	0.0250		0.894
PAC1	0.0076	0.0092	0.0210	0.0339	0.0373	0.0395	0.0546	0.0512		0.955
PAD1	0.0056	0.0070	0.0103	0.0151	0.0160	0.0147	0.0149	0.0147		0.826
PAD2	0.0039	0.0068	0.0141	0.0162	0.0200	0.0225	0.0255	0.0228		0.959
PAE1	0.0046	0.0095	0.0208	0.0352	0.0395	0.0356	0.0389	0.0490		0.928
PAE2	0.0063	0.0082	0.0184	0.0240	0.0269	0.0282	0.0266	0.0293		0.925
PAF1	0.0106	0.0173	0.0239	0.0303	0.0292	0.0379	0.0406	0.0492		0.920
PAF2	0.0032	0.0083	0.0128	0.0155	0.0182	0.0184	0.0152	0.0209		0.847
PAG1	0.0080	0.0100	0.0208	0.0291	0.0394	0.0431	0.0498	0.0423		0.966
Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PBA1	0.0076	0.0104	0.0170	0.0207	0.0246	0.0288	0.0337	0.0449		0.898
PBB1	0.0066	0.0074	0.0191	0.0276	0.0223	0.0206	0.0172	0.0227		0.498
PBB2	0.0009	0.0016	0.0034	0.0035	0.0110	0.0109	0.0160	0.0161		0.891
PBC1	0.0000	0.0000	0.0162	0.0309	0.0373	0.0344	0.0295	0.0388		0.877
PBC2	0.0000	0.0000	0.0000	0.0096	0.0144	0.0147	0.0146	0.0190		0.931
PBD1	0.0053	0.0097	0.0150	0.0147	0.0221	0.0228	0.0231	0.0271		0.955
PBD2	0.0094	0.0162	0.0286	0.0318	0.0426	0.0457	0.0441	0.0538		0.969
PBE1	0.0070	0.0097	0.0156	0.0224	0.0323	0.0373	0.0334	0.0519		0.914
PBE2	0.0007	0.0003	0.0036	0.0069	0.0080	0.0105	0.0106	0.0133		0.973
PBF1	0.0093	0.0122	0.0205	0.0310	0.0368	0.0412	0.0473	0.0459		0.991
PBG1	0.0085	0.0165	0.0203	0.0323	0.0331	0.0345	0.0334	0.0442		0.912
									Average	0.885

In addition to this modest improvement, we note that the Morpheus/MSpC pipeline required significantly less intermediary steps, thus accelerating the data analysis. Some of the additional steps in TPP/ABACUS could be automated from the command-line, but it would likely be a challenge for the average user. Importantly, we found that the Morpheus/MSpC pipeline was faster. Timing tests using the proteasome/*E. coli* spike data generated here showed that the Morpheus/MSpC

Table 2.2: Table of adj_NSAF values for each 20S proteasome subunit (equivalent to dNSAF) generated by analyzing the proteasome spike in experiments with the TPP and ABACUS pipeline. The top half of the table lists α 1-7 (PAA-PAG) where 1 or 2 represent different isoforms) subunits, while the bottom half of the table lists β 1-7 subunits (PBA-PBG where 1 or 2 represent different isoforms)

Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PAA1	125.16	166.81	245.61	355.60	491.76	494.18	504.75	554.97		0.988
PAA2	57.52	70.52	114.60	126.35	132.06	162.25	358.13	154.62		0.493
PAB1	96.13	134.98	207.90	328.16	362.21	416.46	303.60	397.39		0.865
PAB2	29.67	79.42	153.34	282.91	365.22	299.47	310.19	299.93		0.849
PAC1	95.27	112.45	214.24	376.66	389.05	361.39	512.73	488.62		0.925
PAD1	62.58	76.16	106.03	196.03	179.37	182.99	164.04	188.71		0.794
PAD2	39.88	68.33	145.34	178.43	190.84	246.20	243.28	240.01		0.954
PAE1	50.41	97.53	193.18	368.87	464.79	397.27	468.64	534.16		0.951
PAE2	71.38	80.07	184.85	255.98	299.41	290.16	304.77	348.32		0.955
PAF1	132.67	179.60	207.68	253.51	253.04	327.11	334.73	460.76		0.835
PAF2	0.00	0.00	59.62	70.93	111.83	80.29	63.08	65.86		0.606
PAG1	108.09	142.82	271.18	389.94	495.94	530.90	632.72	540.86		0.966
<hr/>										
Ratio	0.091	0.167	0.333	0.500	0.667	0.750	0.800	0.857		Pearsons
PBA1	85.99	113.01	208.38	233.44	273.50	327.23	417.12	560.62		0.851
PBB1	47.77	80.66	195.35	324.53	303.26	220.36	200.51	216.77		0.440
PBB2	17.11	4.54	0.00	0.00	21.05	102.51	144.30	187.86		0.618
PBC1	27.19	40.65	165.46	345.14	364.19	386.59	338.47	373.16		0.885
PBC2	10.64	40.65	0.00	55.58	145.43	158.90	159.74	228.05		0.845
PBD1	50.14	94.46	168.58	167.50	244.25	248.17	252.65	319.61		0.945
PBD2	81.45	149.28	256.88	319.89	390.60	407.02	380.30	490.17		0.950
PBE1	81.01	107.76	196.21	302.90	365.03	404.25	378.67	441.03		0.981
PBE2	0.00	1.52	5.59	8.05	40.06	56.56	59.39	58.56		0.887
PBF1	71.53	112.26	190.45	226.89	259.90	366.60	431.42	420.23		0.943
PBG1	104.93	174.55	232.34	392.37	425.10	397.85	396.42	463.13		0.906
									Average	0.845

pipeline was 1.9-fold faster than the TPP/ABACUS pipeline (Figure 2.5). Such an improvement was expected given that Morpheus completes its searches on average 1.3 to 4.6 times faster than most other search engines available (Wenger and Coon, 2013). Given its simplicity of use, speed, and open source nature, MSpC combined with Morpheus is clearly advantageous over other PSM-based LFQ approaches currently available. Moreover, by being open source, MSpC should allow others to

extend its utility and to serve as a platform for integrating additional open source LFQ approaches into the Morpheus pipeline.

2.3. Methods

2.3.1. Sample Preparation

All chemicals were purchased from Sigma-Aldrich unless otherwise stated. 20S proteasomes were obtained as described previously (Book et al., 2010) from *Arabidopsis thaliana* Col-0 ecotype seedlings in which the 20S proteasome subunit PAG1 ($\alpha 7$) was genetically replaced with a FLAG-tagged variant, with the minor modification of switching to a more stable HEPES buffer during purification. The FLAG peptide used for elution was removed by filtering through an Amicon Ultra 4 10K filter with the elution buffer also exchanged into 8 M urea. Total protein was quantified by the bicinchoninic acid protein assay (Thermo Scientific) using bovine serum albumin as the standard. Approximately 70 μ g of proteasomes were digested overnight at 37°C using a 1:30 trypsin/sample ratio. Peptides were acidified to a final concentration of 1% TFA, desalted on a Waters C18 Sep-Pak containing 50 mg sorbent material, and lyophilized. Total *E. coli* lysates were obtained from Bio-Rad (Cat. 163-2110) with 200 μ g digested as above. Both proteasome and *E. coli* peptides were dissolved in 5% acetonitrile, 95% water, and 0.1 % formic acid. Each MS analysis, performed

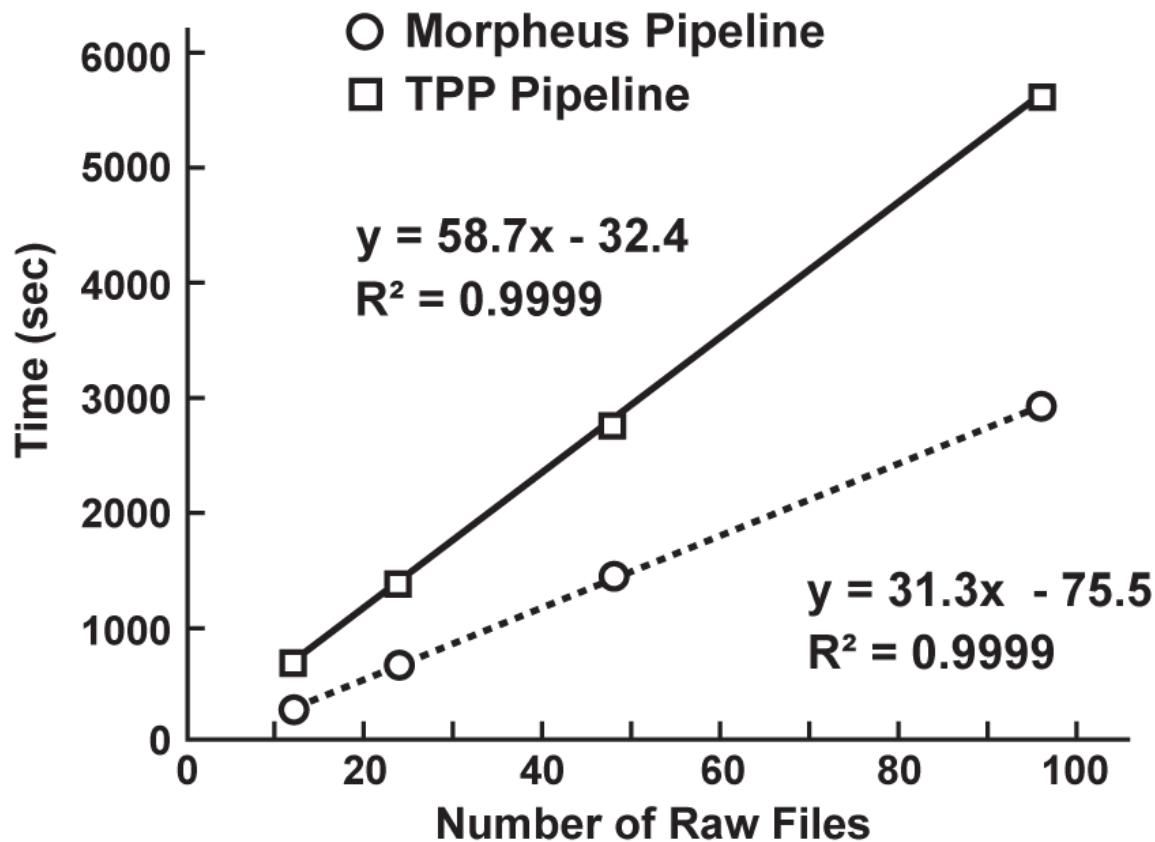


Figure 2.5: MSpC combined with Morpheus works faster than TPP combined with ABACUS. Speed comparisons were performed for 12, 24, 48, and 96 raw MS/MS files generated with the 20S proteasome/*E. coli* lysate samples analyzed in 2.4. On average, Morpheus/MSpC finished the calculations 1.9 times faster than TPP/ABACUS over a ~10-fold range of dataset size.

in triplicae, used 5 μ L volumes prepared with 3, 2, 1.5, 1, 0.5, 0.25, 0.1, or 0.05 μ g of digested proteasomes mixed with 0.5 μ g of digested *E. coli* proteins. This mixtures reflected proteasome/*E. coli* ratios of ~0.091, 0.167, 0.333, 0.500, 0.667, 0.750 0.800, and 0.857, respectively.

2.3.2. Liquid Chromatography and High-Resolution Mass Spectrometry

Samples were analyzed by ultra-high performance liquid chromatography (UPLC) (nanoAcuity, Waters Corporation) connected online to an electrospray ionization LTQ Velos Orbitrap mass spectrometer (Thermo Scientific). Separation employed a 100 x 365 μ m fused silica capillary micro-column packed with 20 cm of 1.7 μ m-diameter, 130- pore size, C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 μ m using a laser puller (Sutter Instruments). Peptides were loaded at a flow-rate of 400 nL/min for 30 min and then eluted over 120 min at a flow-rate of 300 nL/min with a 2% to 30% acetonitrile gradient in 0.1% formic acid. Full-mass scans were performed in the FT Orbitrap with a mass range of 300-1500 m/z at a resolution of 60,000, followed by ten MS/MS high energy C-trap dissociation scans of the ten highest intensity parent ions at 42% normalized collision energy and 7,500 resolution, with a mass range starting at 100 m/z. Dynamic exclusion was enabled with a repeat count of two over the duration of 30 sec and an exclusion

window of 120 sec.

2.3.3. Data Processing

Protein identifications were determined using the Morpheus search engine (Wenger and Coon, 2013). Raw data was searched with the Thermo module of Morpheus revision 151 downloaded and compiled from source code available at <http://sourceforge.net/projects/morpheus-ms/> using Microsoft Visual Studio 2013 professional edition. The following parameters were used to search all databases: unknown precursor charge states - +2, +3, +4; maximum number of MS/MS peaks = 400; assign charge states - enabled; de-isotope - disabled; generate target decoy database on the fly; protease trypsin (no proline rule); maximum missed cleavages = 2; initiator methionine behavior - variable; fixed modification of carbamidomethylation of cysteine; variable modification of oxidation of methionine; maximum variable modification isoforms per peptide = 1024; precursor mass tolerance = ± 2.1 Da monoisotopic (recommended parameters to account for neutral loss); precursor monoisotopic peak correction - disabled; product mass tolerance = ± 0.025 Da monoisotopic; consider modified forms as unique peptides - false; maximum threads = 8; minimize memory usage - false. MSpC quantification of Universal Proteome Standard 2 (UPS2) protein sequences exploited the MS/MS analysis of

individual USP2 proteins (Sigma-Aldrich) mixed at various concentrations with egg or embryo extracts from *Xenopus laevis* available in the PRoteomics IDEntifications (PRIDE) repository (Vizcaino et al., 2013) using identifier - PXD000902 and the available proteomics database - pita_v1.71.protein.name.fa. For our analysis, both the raw MS/MS data and resulting FASTA files for the USP2 and egg and embryo proteomes were obtained from PRIDE. The database used for searching the MS/MS data of *Arabidopsis* proteasomes spiked into total *E. coli* peptides was generated by combining Uniprot K12 *E. coli* reference proteome UP000000625 with a common contaminant database, and then mixing the merged dataset with FASTA sequences for all proteoforms of all known proteasome subunits and associated proteins (Book et al., 2010) obtained from the TAIR10_pep_20101214 FASTA database available within The *Arabidopsis* Information Resource (TAIR) version 10. All FASTA files are available for download in the Supporting Information. The datasets were analyzed by MSpC with a 1% PSM false discovery rate (FDR) and a 1% protein group FDR to determine NSAF, dNSAF, and uNSAF values for each protein group. Two separate analyses were performed in which one unique, or alternatively two unique peptides were required to quantify a protein group. NSAF values were calculated according to formulas 1a, 2a, and 3a from Figure 1 of Zhang et al. (Zhang et al., 2010).

2.3.4. Isoform Incorporation Rates

For the individual subunit analysis of the 20S proteasome the isoform incorporation rates were treated as follows. Given that each proteasome subunit should be incorporated at equal stoichiometry within the 20S particle, we then tested whether the Morpheus/MSpC pipeline could calculate the relative abundance of each subunit and the distribution of isoforms. Here, we divided the dNSAF values for each subunit/isoform by the total number of dNSAF values for the entire complex across all eight total proteasome/*E. coli* lysate ratios tested. This averaged value provided a concentration-independent ratio for the incorporation of each subunit/isoform. We then normalized these values based on a 1/14 stoichiometry of each subunit within the complex to calculate the estimated occupancy versus the expected occupancy of each subunit.

2.3.5. Speed and Accuracy Comparisons of Morpheus/MSpC to the TPP/ABA-CUS Pipeline

The speed and accuracy of MSpC combined with Morpheus was compared to the next most comparable open source software suite for calculating NSAF values; i.e., TPP (Deutsch et al., 2010) combined with ABACUS (Fermin et al., 2011), using the proteasome spike-in experiment files as input. The .raw files were converted to

.mzML files by TPP Build 201411201551-6764 and then searched using the multi-threaded X!Tandem MS/MS search engine (Craig and Beavis, 2004) with the search parameters adjusted as close as possible to that used for the Morpheus searches. A decoy database was generated using the TPP tool DecoyFASTA for use with X!Tandem. Relevant X!Tandem parameters are listed here: parent monoisotopic mass error = \pm 2.1 Da, fragment mass error = \pm 0.025 Da, fixed modifications of carbamidomethylation (57.021464) on cysteine, and variable modification of oxidation (15.994915) on methionine, fully tryptic cleavages, missed cleavage sites = 2 maximum, no refinement and 8 threads. The configuration file used and the test datasets can be found in the Supporting Information. The data were analyzed in the TPP using Peptide Prophet (Keller et al., 2002). Relevant settings are listed: minimum probability = 0.05; minimum peptide length = 7; accurate mass, and nonparametric decoy database to pin down false discovery rate; ignore +1 charged spectra; and run Protein Prophet (Nesvizhskii et al., 2003) after Peptide Prophet. Once completed, all pepxml data from Peptide Prophet contained in a single folder was combined using the command line version of Protein Prophet from the TPP binaries with the following command: ProteinProphet.exe *.pep.xml interact-COMBINED.prot.xml. This post-analysis aggregation was required for running the spectral counting program ABACUS. Here, we note that there are no graphic user interfaces to perform this

post analysis aggregation, which makes this portion of the data analysis more difficult for those unfamiliar with setting up and running programs from the command line. The combined data was analyzed by ABACUS with the following parameters: best peptide probability = 0.99; minimum peptide probability = 0.99; experimental peptide probability = 0; and combined file probability = 0.99 to most accurately match a 1% FDR stringency settings in MSpC. dNSAF values were compared in Microsoft Excel using the CORREL function and squaring the result. Additional timing tests were performed with a subset of the calibration curve data (ratios 0.091, 0.167, 0.333, and 0.500 in triplicate corresponding to 12 .raw files) by increasing file input to 24 (2x), 48 (4x), and 96 (8x) .raw files to determine the time dependence on input size between both pipelines tested (Morpheus/MSpC versus TPP/ABACUS). The timing tests and all data analyses were performed on a computer running Windows 7 Ultimate, with 16 GB of random access memory, and an Intel Core i7-2700k with hyper-threading turned on for eight logical cores.

2.4. Discussion on Requiring Two Unique Peptides to Quantify a Protein

Occasionally, some researchers may want to use a more stringent criterion for quantification such as requiring a protein to have more than one unique identifying peptide. To see how this might affect our data analysis, we re-analyzed our results

shown in Figure 2.3, this time requiring two unique peptides to quantify an individual protein. The results point to a very small increase in linearity observed in the average plots; however, there is a slight decrease in linearity for the egg sample (0.886 to 0.865) and a larger decrease in linearity for the individual UPS2 protein plot for the embryo sample (0.827 to 0.723). The decrease in linearity in the embryo sample is due to the analysis removing a low abundance UPS2 protein (O00762ups) identified with only one unique peptide. While some have suggested that requiring more than one unique peptide to identify a protein is an ideal approach, requiring two peptides for identifications in database searches reduces the number of protein identifications in the target database more than those in the decoy database and results in increased false discovery rates . While we recognize that researchers may want to implement more stringent requirements than what is typically used in database searching to quantify a set of proteins, there are two cases where requiring two unique peptides may not be ideal in a quantitative analysis. Firstly, low abundance proteins that have few PSMs might be identified by only a single peptide and thus be erroneously thrown out of the analysis. Secondly, there may be only one unique peptide that can differentiate between families of homologous proteins. In this second case, requiring two unique peptides would remove these homologous proteins from the MSpC analysis, even if they had a large number of

PSMs. Because of these reasons and because of the decreased linearity observed when requiring proteins to have two unique peptides (Figure 2.3) as compared to one unique peptide (Figure 2.2), we suggest caution in requiring more than one unique peptide per protein.

2.5. Tutorial

Morpheus Spectral Counter (MSpC) Tutorial

Requirements

- 64 bit Windows Installation: For Example Microsoft Windows 7 64 bit, Windows 8 and 8.1 64bit, or Windows 10 64bit
 - This is due to requirements from vendor libraries to process Thermo .raw files directly in this tutorial
 - Download and install proteowizard (<http://proteowizard.sourceforge.net/downloads.shtml>)
 - Installation of proteowizard correctly installs the latest Thermo Vendor .dll's to read .raw files directly
 - Download pre-compiled binaries of revision 151 of the Morpheus Mass Spectrometry Search Engine (Thermo Version)
(https://github.com/dcgemperline/Morpheus_SpC/releases/download/v1.0/Morpheus.Binaries.revision.151.zip), referred from here on out as just Morpheus
 - Download and install MSpC
(https://github.com/dcgemperline/Morpheus_SpC/releases/download/v1.0/MSpC_v1.0.zip)
 - MSpC depends on the latest .NET Runtime that will be installed with MSpC installer if it is not already installed
 - Download the data files from the PRIDE proteomics data repository using the following ID – PXD003002 Data Files (<http://www.ebi.ac.uk/pride/archive/>)
-

Data Analysis Starting from .raw files

1. Once all the requisite software is installed and the Data Files are downloaded you are ready to begin.
2. Unzip **Morpheus Binaries revision 151.zip** into your desired location and start **Morpheus (Thermo).exe** from the Morpheus (Thermo) folder.
3. Once Morpheus starts, verify at the top you are using **Morpheus (Thermo) revision 151**.
4. Add the following .raw files to Morpheus from the Raw Files folder contained in the Data Files folder (multiple files can be selected at once).
 - 04_28_15_A_rep1_c1.raw
 - 04_28_15_A_rep2_c1.raw

- 04_28_15_A_rep3_c1.raw
 - 04_28_15_B_rep1_c1.raw
 - 04_28_15_B_rep2_c1.raw
 - 04_28_15_B_rep3_c1.raw
 - 04_28_15_C_rep1_c1.raw
 - 04_28_15_C_rep2_c1.raw
 - 04_28_15_C_rep3_c1.raw
 - 04_28_15_D_rep1_c1.raw
 - 04_28_15_D_rep2_c1.raw
 - 04_28_15_D_rep3_c1.raw
5. Side Note on Fractionation: If you would like Morpheus, and thus MSpC to output summaries for a set of data by adding up all of the spectra identified in that dataset(as the case may be for fractionation data), place the files you would like to be summarized in seperate folders, such as SampleSet1, Sampleset2, and Morphues will generate protein_groups.tsv and PSMs.tsv specifically for each sampleset. Later in MSpC after selecting the summary.tsv file these will show up as SampleSet1*, SampleSet2*, etc.
 6. Browse for the fasta file contained in the Data Files Folder
uniprot_k12_e_coli_contams_plus_proteasome_and_interactors.fasta.
 7. Verify that Create Target-Decoy Database On The Fly is checked.
 8. Browse for an Output Folder. In this tutorial select browse, leave the default Desktop location highlighted and press Make a New Folder and rename it Morpheus Analysis. Press OK. You should now have a folder on the desktop called Morpheus Analysis, and Morpheus should say you are outputting data to the following output folder, where **NAME** is your username on your machine.
 - C:\Users\NAME\Desktop\Morpheus Analysis
 9. Set the Maximum threads to 2, for a dual core processor, 4 for a quad core processor, 8 for a quad core processor with hyperthreading, and 8 for an 8 core processor. These are reasonable values that will give you decent performance with Morpheus.
 10. Press Search, and the progress bar will indicate search progress for each raw file. On an Intel Core i7 2700K with Morpheus set to use 8 threads, this takes approximately 4.5 minutes.
 11. Close Morpheus if so desired.
 12. On completion of the search open MSpC.
 13. The indicators in Summary Directory Contains Valid Input should be red.
 14. Press the Select the Summary File button and navigate to the summary.tsv file contained in the Morpheus Analysis folder on the Desktop.
 - C:\Users\NAME\Desktop\Morpheus Analysis\summary.tsv
 15. The indicators in Summary Directory Contains Valid Input should now read green indicating that the Morpheus Analysis folder contains all the necessary files from Morpheus to run. (The files required to be in the same directory as summary.tsv are protein_groups.tsv and PSMs.tsv. This will automatically occur if you output all of your Morpheus output into a single output folder.)

16. (Optionally) Select the provided whitelist file contained in the Data Files folder that you downloaded to simplify the output.
 17. Select your desired PSM FDR and Protein FDR (although the default options of 1% are good default options).
 18. Check or uncheck boxes of the desired calculations (NSAF, dNSAF, uNSAF)
 19. (Optionally) Select an Output Directory, otherwise MSpC defaults to the current directory of summary.tsv as an output directory.
 - Let's press Select Output Directory then Make a New Folder here called MSpC Output, press OK.
 20. Press Summarize Data, and the progress bar will indicate the progress of MSpC. This should take ~ 30 seconds or less
 21. The MSpC Output folder will then contain NSAF summaries for each individual .raw file, as well as a summary for all .raw files analyzed. These files are in the tab delimited output format .tsv
-

Data Analysis Starting from previously searched Morpheus Data

1. It is strongly suggested, that you start with the Data Analysis from .raw files tutorial first, and then refer back to the tutorial for previously searched Morpheus data if necessary
2. Open MSpC
3. Select a Morpheus summary.tsv file in a directory that contains the entire output from Morpheus.
4. If all output is there, the indicators in Summary Directory Contains Valid Input should now read green indicating that the folder contains all the necessary files from Morpheus to run MSpC.
5. (Optionally) Select the provided whitelist file contained in the Data Files folder that you downloaded to simplify the output.
6. Select your desired PSM FDR and Protein FDR (although the default options of 1% are good default options).
7. Check or uncheck boxes of the desired calculations (NSAF, dNSAF, uNSAF)
8. (Optionally) Select an Output Directory, otherwise MSpC defaults to the current directory of summary.tsv as an output directory.
 - Let's press Select Output Directory then Make a New Folder here called MSpC Output, press OK.
9. Press Summarize Data, and the progress bar will indicate the progress of MSpC. This should take ~ 30 seconds or less
10. The MSpC Output folder will then contain NSAF summaries for each individual .raw file, as well as a summary for all .raw files analyzed. These files are in the tab delimited output format .tsv

2.6. Acknowledgements

D.C.G. was supported by a grant from the U.S. Department of Energy Office of Science; Office of Basic Energy Sciences; Chemical Sciences, Geosciences, and Biosciences Division (DE-FG02-88ER13968) and a graduate training fellowship from the NIH (5 T32 GM 7133-37). M.S. and L.M.S were supported by a grant from the National Institute of Health/National Institute of General Medical Sciences (1P50HG004942). The authors thank Erin Gemperline, Richard S. Marshall, and Josh Coon for critical reading of the manuscript, and additionally thank Derek Bailey for a critical code review.

2.7. Literature Cited

- Book, A.J., Gladman, N.P., Lee, S.S., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2010). Affinity purification of the Arabidopsis 26S proteasome reveals a diverse array of plant proteolytic complexes. *J. Biol. Chem.* **285**: 25554–25569.
- Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M.** (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* **13**: 2513–2526.
- Craig, R. and Beavis, R.C.** (2004). TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**: 1466–1467.
- Deutsch, E.W., Mendoza, L., Shteynberg, D., Farrah, T., Lam, H., Tasman, N., Sun, Z., Nilsson, E., Pratt, B., Prazen, B., Eng, J.K., Martin, D.B., Nesvizhskii, A.I., and Aebersold, R.** (2010). A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **10**: 1150–1159.

- Fermin, D., Basrur, V., Yocom, A.K., and Nesvizhskii, A.I.** (2011). ABACUS: a computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. *Proteomics* **11**: 1340–1345.
- Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., and Gygi, S.P.** (2003). Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA* **100**: 6940–6945.
- Keller, A., Nesvizhskii, A.I., Kolker, E., and Aebersold, R.** (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74**: 5383–5392.
- Nesvizhskii, A.I., Keller, A., Kolker, E., and Aebersold, R.** (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**: 4646–4658.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M.** (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* **1**: 376–386.
- Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J.** (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**: 1154–1169.
- Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., and Coon, J.J.** (2013). Characterization and quantification of intact 26S proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One* **8**: e58157.
- Thompson, A., Schafer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Johnstone, R., Mohammed, A.K., and Hamon, C.** (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**: 1895–1904.
- Tu, C., Li, J., Sheng, Q., Zhang, M., and Qu, J.** (2014). Systematic assessment of survey scan and MS2-based abundance strategies for label-free quantitative proteomics using high-resolution MS data. *J. Proteome Res.* **13**: 2069–2079.

- Vizcaino, J.A., Cote, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., O'Kelly, G., Schoenegger, A., Ovelleiro, D., Perez-Riverol, Y., Reisinger, F., Rios, D., Wang, R., and Hermjakob, H.** (2013). The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**: D1063–9. Vizcaino, Juan Antonio Cote, Richard G Csordas, Attila Dianes, Jose A Fabregat, Antonio Foster, Joseph M Griss, Johannes Alpi, Emanuele Birim, Melih Contell, Javier O'Kelly, Gavin Schoenegger, Andreas Ovelleiro, David Perez-Riverol, Yasset Reisinger, Florian Rios, Daniel Wang, Rui Hermjakob, Henning eng BB/I024204/1/Biotechnology and Biological Sciences Research Council/United Kingdom WT085949MA/Wellcome Trust/United Kingdom Research Support, Non-U.S. Gov't England 2012/12/04 06:00 Nucleic Acids Res. 2013 Jan;41(Database issue):D1063-9. doi: 10.1093/nar/gks1262. Epub 2012 Nov 29.
- Wenger, C.D. and Coon, J.J.** (2013). A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *J. Proteome Res.* **12**: 1377–1386.
- Wong, J.W. and Cagney, G.** (2010). An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods Mol. Biol.* **604**: 273–283.
- Yang, P., Fu, H., Walker, J., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D.** (2004). Purification of the *Arabidopsis* 26S proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *J. Biol. Chem.* **279**: 6401–6413.
- Zhang, B., VerBerkmoes, N.C., Langston, M.A., Uberbacher, E., Hettich, R.L., and Samatova, N.F.** (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* **5**: 2909–2918.
- Zhang, Y., Wen, Z., Washburn, M.P., and Florens, L.** (2010). Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. *Anal. Chem.* **82**: 2272–2281.
- Zybailov, B., Mosley, A.L., Sardiu, M.E., Coleman, M.K., Florens, L., and Washburn, M.P.** (2006). Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* **5**: 2339–2347.

Chapter 3

PROTEOMIC CHARACTERIZATION OF AFFINITY-PURIFIED ARABIDOPSIS PROTEASOME PARTICLES IDENTIFIES SUB-COMPLEX SPECIFIC INTERACTORS

3.1. Statement of Attribution

Initial mutant characterization of rpt4 mutants was performed by Drs. Kwang-Hee Lee and Richard S. Marshall, postdoctoral fellows in the Vierstra Lab. Additionally, Dr. Lee generated the 2XFLAG-RPT4b transgenic line in the rpt4b-2 background. I created the 2XFLAG-RPT4a rpt4a-1 transgenic lines. I developed the affinity purification via the RP with minor modifications from (Book et al., 2010) by changing the primary buffer to HEPES. I performed all the mass-spectrometry sample preparation and data analysis in this chapter. The mass spectrometry was either performed by Dr. Mark Scalf in Dr. Lloyd M. Smiths lab (Department of Chemistry, University of Wisconsin Madison), or by Dr. Fionn McLoughlin in the Vierstra lab at Washington University. I cloned the CP interactors PAP1, PBAC1, PBAC2, PBAC3, and PBAC4, and performed yeast-two-hybrid (Y2H) analysis on these. While the streaking and yeast transformations were performed by myself, Dr. Marshall per-

formed the final spotting of the yeast-two hybrid. All BiFC data was generated by myself. The HA-PAP1 transgenic line was created by Dr. Adam Book. I developed the affinity purification for HA-PAP1 and performed the immunoblot analyses. Native PAGE was performed by Dr. Marshall.

We intend to submit this chapter to J. Biol. Chem. for publication as a research article in the near future. A suggested list of authors is as follows: David Gemperline, Richard S. Marshall, Kwang-Hee Lee, Fionn McLoughlin, Mark Scalf, Lloyd M. Smith, and Richard D. Vierstra.

3.2. Summary

The 26S proteasome is an ATP-dependent protease complex responsible for the selective removal of numerous eukaryotic proteins following their modification with polyubiquitin chains. It consists of two sub-complexes, a self-compartmentalized core protease (CP) that degrades polypeptides within a central chamber, and a regulatory particle (RP) that caps one or both ends of the CP and is responsible for recognizing and unfolding ubiquitylated substrates, and transporting them into the CP lumen. Previous mass spectrometric (MS) analyses of *Arabidopsis* 26S proteasomes affinity-purified via the CP subunit PAG1 (7), identified a complex assortment of particles that are assembled using paired isoforms for many of the

33 core subunits, as well as a small assortment of associated factors that might help in assembly, regulation, substrate specificity, and ubiquitin recycling. Here, we extended these studies by affinity purifying *Arabidopsis* 26S proteasomes with either tagged isoform of the RP subunit RPT4 (named RPT4a and RPT4b). Label-free quantitative (LFQ) MS analyses of the preparations detected both isoforms for many of the other CP and RP subunits, suggesting that plants incorporate subunit isoforms randomly into the holo-protease. We also detected an array of interacting factors, including likely orthologs of the yeast assembly chaperones Ump1 and Pba1-4 for the CP, and Nas2, Nas6, and Hsm3 for the RP. We additionally identified a novel plant-specific protein bound to the CP that interacts with the putative *Arabidopsis* Pba1 ortholog, possibly contributing to CP assembly. By coupling native PAGE and MS analysis with proteasome preparations isolated from proteasome inhibitor-treated seedlings, we identified several assembly intermediates also harboring assembly chaperones and the proteasome regulators PA200 and ECM29. Taken together these data suggest that *Arabidopsis*, like yeast, may dynamically employ a suite of assembly factors to construct the mature 26S particle.

3.3. Introduction

The selective removal of proteins that are non-functional or have key regulatory roles is essential to normal plant growth and development. A major pathway for this selective removal is the ubiquitin 26S proteasome system, in which exposed lysine residues on target-proteins are post translationally modified with the 76-amino-acid polypeptide ubiquitin and subsequently degraded by the central enzymatic effector in this process, the 26S proteasome (reviewed in (Finley, 2009; Livneh et al., 2016; Vierstra, 2009)). The 26S proteasome consists of two major sub-complexes, the core protease (CP), which is responsible for degrading the substrate, and the regulatory particle (RP), which caps the CP at one or both ends and is responsible for recognizing and unfolding ubiquitylated substrates. Compositionally, the CP is made up of 28 subunits configured into four stacked hetero-heptameric rings made up of 7 subunits in an 1-7/1-7/1-7/1-7 arrangement. The N-termini of the 1, 2, and 5 subunits are enclosed in a central chamber by the surrounding 21 subunits, and together they provide peptidylglutamyl-peptide-hydrolyzing activity (1), trypsin-like (2), and chymotrypsin-like (5) activities. The 7 subunits possess N-Terminal extensions that gate entry to the chamber, forming an axial pore that is only opened when substrates are actively processed by an associated RP, or when activated by alternative capping particles such as PA200 (Dange et al., 2011; Sadre-Bazzaz et al.,

2010).

The RP consists of two major sub-classes of proteins including six related AAA-ATPase (RPT) subunits (RPT1-6) and at least 15 non-AAA-ATPase (RPN) subunits (RPN1-3, 5-13, and SEM1) (Finley, 2009; Paraskevopoulos et al., 2014; Russell et al., 2013). The RPT subunits form a six-membered RPT ring responsible for unfolding the polypeptide substrate in an ATP dependent manner, and channeling the substrate through the axial pore into the central CP chamber. The RPN subunits have several differing functions including binding ubiquitylated substrates (RPN1, RPN10, RPN13, and SEM1 (Elsasser et al., 2004; Paraskevopoulos et al., 2014; Schreiner et al., 2008)), binding ubiquitin shuttle proteins such as DSK2 and RAD23 (Elsasser et al., 2002; Farmer et al., 2010; Fatimababy et al., 2010; Lin et al., 2011), and releasing ubiquitin bound to substrates through the de-ubiquitylating activity of RPN11 (Verma et al., 2002; Yao and Cohen, 2002). RPN5-7 and RPN9 form a glove-like structure around the RPT ring providing structural support for adjacent subunits (Lander et al., 2012; Lasker et al., 2012; Unverdorben et al., 2014), with RPN6 acting as a molecular clamp that holds the RP and CP together (Pathare et al., 2012).

Over the last several years, analyses of the plant 26S proteasome have also revealed some insights into its composition. Genetic analyses of the *Arabidopsis*

genome showed that plants typically encode two paralogs for most proteasome subunits (Fu et al., 1998). Almost all of these paralogs were identified in chromatographic preparations of the Arabidopsis proteasome (Yang et al., 2004), and in affinity preparations that exploited a FLAG tagged CP subunit PAG1 (Book et al., 2010). Like in Arabidopsis, duplication of RPT subunits have also occurred in monocots, with rice having duplicated subunits for RPT1, 2, 4 and 5, suggesting that there may be some evolutionary benefit to having multiple copies of the RPT family (Shibahara et al., 2004). Taken together these data suggest that plants have the capacity to generate a highly diverse set of proteasome isotypes by preferentially incorporating specific subsets of isoforms. Precedents for proteasome isotypes exist in mammals, which form the alternative immunoproteasome and thymoproteasome isotypes by incorporating alternative isoforms of the catalytic subunits (Murata et al., 2007; Nandi et al., 1996). Similarly, the mammalian testes proteasome isotype incorporates an alternative 4 isoform (Belote et al., 1998). In plants, genetic evidence suggests that some subunit paralogs may have distinct functions. For instance, RPN1a and RPN1b play differing roles in embryogenesis (Brukhin et al., 2005). Furthermore, overexpression of RPN5a induces an early senescent phenotype while overexpression of RPN5b does not (Book et al., 2009). Despite these observations, it remains unclear if plants assemble their proteasome

subunit isoforms into compositionally and functionally distinct isotypes.

Outside of the core subunits, other proteins are known to associate with the proteasome transiently or sub-stoichiometrically. These proteasome-associated proteins, or PAPs, perform a variety of functions. DSK2 and RAD23 contribute to the shuttling of ubiquitylated substrates to the proteasome (Farmer et al., 2010; Fatimababy et al., 2010; Lin et al., 2011), RPN13 helps bind ubiquitylated targets (Schreiner et al., 2008), UCH37 and other de-ubiquitylating enzymes (DUBs) (Van-derLinden et al., 2015) help process substrates bound to the RP, and assembly chaperones assist in the construction of both the CP and RP. While the process of plant proteasome assembly is largely unknown, a model has been established in yeast and mammals (see figure intro x). In both systems, the Pba3 chaperone (PAC3 in mammals) interacts with Pba4 (PAC4), forming a Pba(PAC)3/4 heterodimer that contacts the 5 subunit, and likely adjacent subunits, aiding in early stages of CP assembly (Kunjappu and Hochstrasser, 2014; Yashiroda et al., 2008). Similarly, Pba1 (PAC1) interacts with Pba2 (PAC2) to form a Pba(PAC)1/2 heterodimer that, together with Ump1, helps to form 15S half-barrels consisting of 1-7/1-6 notably lacking the 7 subunit (Kunjappu and Hochstrasser, 2014; Marques et al., 2007). CP maturation occurs concomitant with autolytic cleavage of subunit propeptide-appendages exposing proteolytic N-terminal threonines and subsequent degradation of Ump1,

the first substrate of the CP (Ramos et al., 1998). In addition to these dedicated CP assembly chaperones, in humans and yeast a different suite of other chaperones also guides assembly of the RP, including Nas2, Nas6, Hsm3, and Rpn14 (see figure intro x). These RP assembly chaperones typically bind to the C-terminal domains of specific RPT subunits forming modules (Nas2 with Rpt4 and 5, Nas6 with Rpt3, Hsm3 with Rpt1,2 and Rpn1, and Rpn14 with Rpt6) that, importantly, prevent the assembling RPT ring from forming a pre-mature CP-RP interface (Park et al., 2010). In plants, prior MS/MS analyses of Arabidopsis proteasomes identified only one putative CP assembly chaperone, PBAC2 (the nomenclature used here is different from yeast (Pba) and mammals (PAC) as both are standard nomenclature for the 1 and 1 subunits in Arabidopsis), which is an ortholog to PAC2, suggesting that other CP assembly chaperones remain to be identified (Book et al., 2010). Additionally, these proteomic analyses failed to identify obvious plant RP assembly orthologs, such as NAS2, NAS6 and HSM3, despite the identification of these orthologs in the Arabidopsis genome (Book et al., 2010). Taken together, the current data suggest that our understanding of plant PAPs in general, and CP and RP assembly chaperones in particular, remains incomplete.

While tools exist in other organisms to affinity-purify the 26S proteasome via the RP, including Pro-A tagged variants of RPN11, and RPT1 in yeast (Leggett et al., 2005,

2002), and RAD23-based strategies that bind RPN1 in mammals (Besche et al., 2009), such systems were not available in plants until this work. We also reasoned that our existing CP-based affinity purification might miss some RP-specific PAPSs. To overcome this challenge and to explore the possibility of plants forming proteasome isotypes, we sought to develop an RP affinity-purification strategy using two FLAG epitope-tagged RP subunit isoforms, RPT4a and RPT4b. RPT4 was specifically chosen for affinity tagging as prior structural analyses indicated that its N-terminus was likely solvent exposed (PDB 4CR2) (Beck et al., 2012). We also exploited the ATP dependence of the interaction between the CP and RP to enrich for both of these sub-complexes specifically (Book et al., 2010; Liu et al., 2006), enabling us to identify novel CP and RP interacting partners, including several putative assembly chaperones. Follow-up interaction studies suggest a role for several of these associated proteins in CP assembly. Finally, proteomic analyses of native-gel separated complexes from tissues treated with the potent and specific proteasome inhibitor MG132 identified novel sub-complexes that contain distinct subgroups of these putative assembly chaperones. From these results, we propose a model that assigns a putative role for several of these novel associated proteins in distinct phases of proteasome assembly.

3.4. Experimental Procedures

3.4.1. Transgenic Plants and Growth Conditions

The rpt4a-1, rpt4a-3, rpt4a-4, and rpt4b-1, rpt4b-2, and rpt4b-3 T(transfer)-DNA insertion mutants in the *Arabidopsis thaliana* Columbia-0 ecotype (Col-0) (SALK_052372, SALK_128087C, SALK_135246, SALK_108556, SALK_108557, and SALK_101982C, respectively) were obtained from the *Arabidopsis* Biological Resource Center (The Ohio State University, Columbus, OH). All mutants were backcrossed three times to the Col-0 parent and then made homozygous by self-fertilization. The PAG1::PAG1-FLAG pag1-1 material was as previously described (Book et al., 2010). The rpt4a and rpt4b alleles were tracked by genomic PCR using the T-DNA-specific left border primer (Lba1) in combination with gene-specific primers (see supplemental Table S1 for all primers used in this study). The exact T-DNA insertion sites were determined by sequencing the T-DNA-specific PCR products using BigDye Terminator Sequencing v3.1 (University of Wisconsin - Madison Biotechnology Center). For RT-PCR, total RNA was extracted from 7-day old liquid grown seedlings of the indicated genotypes using the RNAeasy plant mini kit (Qiagen), and then converted to cDNA using oligo-(dT)20 primers and the SuperScript III first-strand synthesis system (Invitrogen) both as according to manufactures instructions. RT-PCR of the

converted cDNA was then performed with primer pairs (P1-P9).

Transgenic plants expressing FLAG-tagged variants of RPT4a and RPT4b were constructed as follows. The genomic region including the full coding sequence of RPT4a (AT5G43010), the 600bp sequence upstream of the ATG start codon, which included the 5-untranslated region (UTR), and the 3-UTR, was PCR-amplified from Col-0 genomic DNA using primer pair (P10 and P11) and cloned into the pDONR221 (Invitrogen) vector via the Gateway BP Clonase II reaction (Invitrogen). Two FLAG (DYKDDDDK) tags were inserted using two rounds of site directed mutagenesis (Edelheit et al., 2009) with primer pairs P12, P13 and P14, P15. The sequence-confirmed 2xFLAG-RPT4a clone was recombined into the plant transformation vector pMDC123 (which encodes resistance to the herbicide Basta (glufosinate)) via the Gateway LR Clonase II reaction (Invitrogen). The genomic region including the full coding sequence of RPT4b (AT1G45000) (from start to stop codon, lacking the 3UTR), plus the 2kb sequence upstream of the ATG start codon (including 5 UTR) was PCR-amplified from Col-0 genomic DNA using primer pairs P16, P17 (containing the 2XFLAG tag, NcoI, and SmaI digestion sites) and P18, P19 (containing a PstI and NcoI sites) for the coding sequence and upstream region, respectively. The amplified regions and the recipient pCAMBIA3301 vector were digested with NcoI, SmaI, and PstI (New England Biolabs (NEB)), and ligated

together with T4 DNA ligase (NEB). The recombinant plasmid was cloned into the Escherichia coli TOP10 strain (Invitrogen), and confirmed by sequencing.

The PAP1 coding sequence was PCR-amplified from Col-0 cDNA using primer pair P20, P21, and cloned into the pDONR221 vector via the Gateway BP Clonase II reaction. The PAP1 coding sequence was then recombined in-frame into the plant transformation vector pMDC99 which harbored a 3x-HA tag upstream of the recombination site (resulting in a 3x-HA tag, followed by a short linker RSS- RGVHHMITTLYTKVEMK, followed by the initiator methionine for PAP1) driven by a UBQ10 promoter. This was accompanied by a hygromycin B resistance gene. All constructs were transformed into plants by the Agrobacterium tumefaciens-mediated floral-dip method with the GV3101 agrobacterium strain (Gelvin, 2003; Zhang et al., 2006). The constructs for RPT4a::FLAG-RPT4a and RPT4b::FLAG-RPT4b were transformed into plants harboring the rpt4a-1 and rpt4b-2 insertions, while the construct for UBQ10::HA-PAP1 was transformed into PAG1-FLAG pag1-1 plants. T1 plants were selected with the relevant plant resistance markers (Basta for both the RPT4a pCAMBIA3301 vector and the RPT4b pMDC123 vector, and hygromycin for the PAP1 pMDC99 vector). T2, and T3 plants were obtained by self-pollination. T3 progeny analysis of herbicide/antibiotic-resistant segregation was used to identify homozygous T2 plants. Expression of the RPT4a/b trans-

genes was confirmed by SDS-PAGE followed by immunoblot analysis with both M2 anti-FLAG antibodies (Sigma Catalogue Number (Cat. No.) F1804) and anti-RPT4 antibodies generated by Kwang-Hee Lee (University of Wisconsin) from proteins extracted from 7-day old seedlings grown on 0.8% agar plates containing standard plant growth medium (GM medium, containing MS salts, and Gamborgs B5 vitamins (Gamborg et al., 1968; Julio and Jose, 2006; Murashige and Skoog, 1962)), and 2% sucrose). Expression of HA-PAP1 was confirmed by SDS-PAGE followed by immunoblot analysis with anti-HA antibodies (Sigma anti-HA antibody, Cat. No. H6908) from proteins extracted from 7-day old seedlings grown on 0.8% agar plates containing GM medium, and 2% sucrose.

3.4.2. Sequence and Phylogenetic Analyses

Plant protein sequences were obtained using the Basic Local Alignment Search Tool (BLAST) of *Arabidopsis* PAP1 and PBAC1-4 orthologs, against their respective plant genomes from Phytozome v8 . *Saccharomyces cerevisiae* and *Homo sapiens* Pba/PAC1-4 identifiers were obtained from literature searches of the relevant proteins. The plant PBAC nomenclature is used for consistency. Related animal, and yeast sequences were obtained using top hits in each respective proteome from iterative PSI-BLAST searches using the BLOSUM62 substitution matrix, and

hits above the default expectation value (E-Value) threshold of 0.005 in each iteration (Altschul et al., 1997). Sequence alignments and identity matrices were determined using the Clustal Omega, which selects the appropriate Gonnet 40, 80, 120, 160, 250 and 350 protein substitution matrices in its alignment algorithm (Gonnet et al., 1992; Sievers and Higgins, 2014; Sievers et al., 2011). Alignments were visualized with BOXSHADE (v3.2.3, <https://sourceforge.net/projects/boxshade/>). MrBayes (v3.2.2) was used to generate the phylogenetic trees for 1 million generations with a relative burn-in of 25% (Ronquist et al., 2012). The analysis was run until convergence, and the resulting consensus tree was visualized in FigTree (v1.4.2, <http://tree.bio.ed.ac.uk/software/figtree/>), with final figures made in Adobe Illustrator CS6.

3.4.3. Proteasome Affinity Purifications using PAG1-FLAG and FLAG-RPT4a/b

Affinity purification of proteasomes was performed as described previously with a minor modification in the primary buffering agent, which was switched from Tris to HEPES, as it has a more stable pKa (Book et al., 2010). Liquid cultures were grown for 10 days under constant 150 µmol/m²/s light in GM medium with 2% sucrose at 21-23°C with gentle shaking (90 rpm). For MG132 treatments, at day 9, the growth medium was replaced with fresh medium containing 50 µM MG132

(SelleckChem) in DMSO for the treated samples, or fresh medium with equivalent volumes of DMSO lacking MG132 for the untreated samples, and harvested after 16 hours. 2.5 - 5 g of seedlings were pulverized by mortar and pestle at liquid nitrogen temperatures, extracted at 4°C with 1.25 mL/g of fresh weight of buffer A.1 (50 mM HEPES pH7.5, 50 mM NaCl, 10 mM MgCl₂, and 10% (v/v) glycerol, 2 µM chymostatin, 2mM phenylmethane sulfonyl fluoride (PMSF), 5mM dithiothreitol (DTT), with or without 20 mM ATP), filtered through Miracloth (EMD Millipore), and clarified at 30,000 × g for 30 min. Clarified protein extracts were applied to 50 µL of M2 anti-flag affinity resin (Sigma, Cat. No. A2220) (pre-equilibrated by washing with 1 mL buffer A.1 three times), washed three times with 40 column volumes (2 mL) of buffer A.1. Samples were eluted in approximately 250 µL of 500 ng/µL FLAG peptide (DYKDDDDK synthesized at the University of Wisconsin Biotech Center, Madison WI) in A.1 buffer without PMSF and chymostatin. Samples were stored at -80°C until use.

3.4.4. SDS-PAGE and Native-PAGE Analysis and Silver Staining

SDS-PAGE utilized a Tris-glycine buffering system and was prepared according to the protocol originally established by Laemmli (Laemmli, 1970) and were performed as described in . Gels were made using 40% (w/v) acrylamide/bis-acrylamide

solution (29:1, Bio-Rad), SDS-PAGE resolving buffer (375 mM Tris-HCl pH 8.8 and 0.125% SDS), SDS-PAGE stacking buffer (125 mM Tris-HCl pH 6.8 and 0.125% SDS), 0.1% ammonium persulphate (APS) and polymerized with N,N,N,N-tetramethylethylenediamine (TEMED) with a final acrylamide concentration of 11% and 3.5% for the resolving and stacking gels, respectively. The gels were cast using a Hoeffer Gel casting system into 18 x 16 cm glass plates using 0.75 mm spacers. Proteins were separated with a Hoeffer SE600 electrophoresis unit using Tris-glycine SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS, pH 8.3; Bio-Rad) at a constant current of 15 mA for 6 hours. Approximately 32 µL out of 250 µL of each affinity-purified proteasome preparation was loaded for SDS-PAGE with Laemmli sample buffer.

Native-PAGE gels were prepared according to Book et al. (Book et al., 2010), and performed exactly as described in . A discontinuous Native-PAGE system was used. The lower gel contained 4.6% acrylamide (w/v), 0.12% Bis (w/v), 2.3% sucrose (w/v), Tris-Borate-EDTA (TBE) buffer (89 mM Tris, 89 mM boric acid, 2mM EDTA, pH8.3, Bio-Rad), 0.1% APS, 5mM MgCl₂, 1 mM ATP. The upper gel contained 2.5% acrylamide (w/v), 0.62% Bis (w/v), 2.3% sucrose, TBE, 0.1% APS, 5mM MgCl₂, 1mM ATP. Gels were cast and run as described above. Gels were run using a Hoeffer SE600 electrophoresis unit with TBE buffer supplemented with 0.5 mM ATP instead of Tris-glycine SDS running buffer. Approximately 32

μ L of each proteasome affinity purification was supplemented with 5 μ L of 0.03% (w/v) xylene cyanol to visualize loading. Gels were run at 4°C at 50V for 16 or 32 (extended Native-PAGE) hours. Silver-staining of all gels was accomplished using a mass-spec safe short silver nitrate staining protocol by Chevallet et al .

3.4.5. MS/MS Sample Preparation

All chemicals for MS/MS analyses were obtained from Sigma-Aldrich, unless otherwise noted. All solvents used were obtained from Thermo-Fisher and were LC/MS grade or J.T Baker® grade unless otherwise noted. For in-solution digests, 200 μ L of eluent from each proteasome affinity purification was dried down in a Speed Vac centrifugal evaporator (Savant Instruments, model number SUC100H) until approximately 25 μ L of volume remained, then resuspended in 100 μ L of 8M urea. The samples were reduced with 10 mM DTT (1 hour RT), alkylated with 40 mM iodoacetamide (1 hour RT in the dark), quenched with 40 mM DTT (10 min RT) and then digested with approximately 1 μ g of trypsin overnight at 37°C. The samples were acidified with 10% trifluoroacetic acid (TFA) until the pH was 2, desalted on C18 Bond Elut OMIX Tips (Agilent) as according to manufacturers instructions, and eluted in 75% acetonitrile (ACN), 0.1% acetic acid. The samples were then dried in a Speed Vac centrifugal evaporator and resuspended in 30 μ L of

5% acetonitrile, 0.1% formic acid and stored at -80 oC until analyzed by MS/MS.

For digestion of samples run on a native-PAGE gel, silver-stained protein bands were cut into 1-mm³ pieces with a clean razorblade, washed with MilliQ (EMD Millipore) water, and destained with a 1:1 solution of freshly prepared 100 mM sodium thiosulfate (Na₂S₂O₃) and 30 mM potassium ferricyanide (K₃Fe(CN)₆) until colorless. Silver-ions were then removed from the gel slices by washing twice with MilliQ water. The slices were dehydrated with a 50% ACN in 25 mM ammonium bicarbonate (NH₄HCO₃) solution, then dried in a Speed Vac. They were rehydrated with a reducing solution (25 mM DTT in 50 mM ammonium bicarbonate) for 20 minutes at 56oC, to cleave disulfide bonds. The gel pieces were then treated with an alkylating solution (55 mM iodoacetamide in 50 mM ammonium bicarbonate) in the dark for 20 minutes at RT to alkylate free cysteines. The alkylating solution was then removed, the gel slices were washed with MilliQ water, dehydrated in 50% ACN in 25 mM ammonium bicarbonate solution, and then dried in a Speed Vac. The ProteaseMax (Promega) in-gel digest protocol was then used to aide in digestion and peptide extraction from gel slices (Saveliev et al., 2013). Samples were then desalted with self-packed C18 stage-tips (Rappaport et al., 2003) using a spin protocol (Yu et al., 2014), dried, and resuspended as described above.

3.4.6. MS/MS Analyses

In-solution-digested peptide mixtures were separated using a Dionex U3000 nano-flow ultra-high performance liquid chromatography system (nUHPLC, Thermo Scientific). The peptides were separated using an Acclaim® PepMap RSLC C18 column, 2 m particle size, 100 Å pore size, 75 m x 15 cm (Thermo Scientific) using a 90 minute linear gradient from 5% ACN to 33% ACN with 0.1% formic acid. Electrospray ionization (ESI) of separated peptides was performed with a NanoSpray Flex ion source and ionized peptides were subsequently analyzed with a Thermo Q-Exactive Plus high-resolution accurate-mass orbitrap MS/MS. A data-dependent acquisition program was run with an MS1 mass range of 380-1500 m/z at 70,000 mass resolution, with an automatic gain control (AGC) target of 1e6 ion intensity. The top 15 peaks were chosen for higher energy collisional dissociation (HCD) fragmentation at 28% normalized collision energy with a dynamic exclusion set to 10s. Product MS2 scans were acquired with a 200-2000 m/z mass range at 17,500 mass resolution, with an AGC target of 8e3. All peaks were recorded in profile mode. All raw files associated with these experiments were deposited in the PRoteomics IDEntification (PRIDE) database with identifier XXXXXXXX.

Peptide samples obtained from native-gel slices were instead analyzed using a nUHPLC (nanoAcuity, Waters Corporation) connected online to an electrospray

ionization LTQ Velos Orbitrap mass spectrometer (Thermo Scientific). Separation of the peptide mixture employed a 100 x 365 μm fused silica capillary micro-column packed with 20 cm of 1.7 μm -diameter, 130- \AA pore size, C18 beads (Waters BEH), with an emitter tip pulled to approximately 1 μm using a laser puller (Sutter Instruments). Peptides were separated with a 2% to 30% acetonitrile gradient in 0.1% formic acid. Full-mass scans were performed in the FT Orbitrap with a mass range of 300-1500 m/z at a resolution of 60,000, followed by ten MS/MS high energy C-trap dissociation scans of the ten highest intensity parent ions at 42% normalized collision energy and 7,500 resolution, with a mass range starting at 100 m/z. Dynamic exclusion was enabled with a repeat count of two over the duration of 30 sec and an exclusion window of 120 sec.

3.4.7. MS/MS Data Processing

For the proteasome associated protein studies, raw spectra files were processed with MaxQuant version 1.5.3.30 (Cox and Mann, 2008) using the Andromeda search engine against a TAIR10_pep_20101214_updated.fasta file obtained from The Arabidopsis Information Resource (TAIR) version 10 . Andromeda searches were performed with MS1 mass tolerance of 20ppm and an MS2 mass tolerance of 10ppm. Up to two missed cleavages were allowed, along with the fixed modifications of

carbamidomethyl on cysteine, and variable modifications including oxidation of methionine, acetylation of protein N-termini, and GlyGly (ubiquitylation remnant after cleavage with trypsin) on lysines for possible ubiquitylation events, were specified. Peptides and protein groups were identified with a 0.01 false discovery rate (FDR).

Label-free-quantitative analyses were performed using MaxQuants MaxLFQ (Cox et al., 2014) algorithm (generating the MaxLFQ values output) with default settings except that only unique peptides were used for quantification so that protein isoforms could be discriminated. Matching was set to match from and to samples to increase the number of quantifiable proteins per run, using the 'MaxQuants match between runs' feature. MaxQuants proteinGroups.txt file was then processed in Perseus (Tyanova et al., 2016), a software suite for analyzing mass spectrometry data. Ten proteins were marked as likely contaminants including nitrilases, cruciferins, and seed storage albumins (full list available in Table S2). MaxLFQ values were averaged for both technical replicates, and missing MaxLFQ quantification values were imputed with a 1.8 fold downshifted normal distribution with width 0.3. Volcano plots were generated comparing each proteasome affinity purification against its respective wild-type control using Perseuss volcano plot function using a two-sided t-test, 0.01 FDR, S0 value of 2, and 250 permutations. The resulting

volcano plots were plotted using the Seaborn python graphing library (Tyanova et al., 2016; Waskom, 2016), and edited in Adobe Illustrator. Proteins that were called by Perseus as statistically significant were included in tables (x through x). KDE estimations for CP and RP subunits were performed using the Seaborn graphing library, with the median test and Kolmogorov-Smirnov (KS) tests for enrichment performed using `scipy.stats`. For the bar graphs in Figure 3, missing values were imputed with a Log₂(MaxLFQ value) of 15 so that a zero value could be more easily viewed in log scale.

For the MS/MS analyses of native gel slices the data were searched with the Morpheus search engine (Wenger and Coon, 2013) against same protein database used above. The recommended settings of MS1 search tolerance of 2.1 Da, MS2 search tolerance of 0.01 Da, and 1% maximum FDR were used. Fixed modifications of carbamidomethyl on cysteine, and variable modifications of oxidation on methionine, and GlyGly on lysine with up to two missed cleavages were used. LFQ was completed with Morpheus Spectral Counter (Gemperline et al., 2016) to calculate dNSAF values as described in Chapter 2. The data were plotted in R using the `heatmap.2` function, with centroided hierarchical clustering based on Pearson's correlation as the distance function.

3.4.8. Yeast-two hybrid, bimolecular fluorescent complementation, and HA-PAP1 immunopurification

Assays for protein-protein interactions between PBAC1-4 and PAP1 were performed using the ProQuest® two-hybrid system which uses the pDEST22 and pDEST32 vectors, and the *Saccharomyces cerevisiae* MaV203 strain all available from Thermo Fisher Scientific. PBAC1-4 and PAP1 were amplified from *Arabidopsis* cDNA with primers (P20 P29) and transformed into the pDONR221 entry vector (Thermo Fisher Scientific). Genes were then recombined using Gateway LR Clonase II (Thermo Fisher Scientific) into vectors pDEST22 and pDEST32. Yeast strain MaV203 was transformed with pairwise gene combinations in pDEST22 and pDEST32, or the empty vector controls. Interactions were tested by growing transformants for 2 days at 30°C on synthetic complete medium lacking leucine, tryptophan, and histidine, with added 30mM 3-amino-1,2,4-triazole.

For BiFC assays, PBAC1-3 and PAP1 cDNA sequences in pDONR221 were recombined using Gateway LR Clonase II (Thermo Fisher Scientific) into either pSITE-N-EYFP-C1 or pSITE-C-EYFP-C1 vectors (ABRC stocks CD3-1648 for NYFP or CD3-1649 for cYFP respectively), and subsequently transformed into *Agrobacterium tumefaciens* strain GV3101 (Gelvin, 2003). Overnight cultures were diluted in resuspension buffer (10 mM MgCl₂, 10 mM MES pH 5.7) to a final OD600 of 0.5

and then infiltrated by syringe into 4-6 week old Nicotiana benthamiana leaves. After 36 h, fluorescence in the infiltrated regions was visualized using a Zeiss 510 Meta confocal laser scanning microscope, and images were analyzed with Zeiss LSM image browser v3.5 and ImageJ.

The immunopurification of PAP1 to determine if PAP1 could co-purify with 26S proteasome subunits was performed as follows. HA-PAP1-expressing tissue in the PAG1-FLAG pag1-1 background were grown in liquid culture as described previously (Book et al., 2010). Approximately 2g of tissue were ground using liquid nitrogen chilled mortar and pestles, and soluble proteins were extracted in 2.5 mL of buffer A.1. 25 µL of HA-EZview (Sigma-Aldrich) beads were incubated overnight with 1.75 mL of protein extracts in 2mL Eppendorf tubes at 4°C with constant rotation. The HA beads were washed with 3 x 1mL of buffer A.1, and then eluted with 100µL of 2x protein sample buffer (Laemmli (Laemmli, 1970)).

3.5. Results

3.5.1. **Arabidopsis Proteasomes Can Be Effectively Affinity-Purified Using Epitope-Tagged Versions of Either Isoform of the Regulatory Particle Subunit RPT4**

As a first step toward purifying *Arabidopsis* 26S proteasome via the RP, and exploring the possibility that plants assemble unique proteasome isotypes, we sought out candidate subunits amenable to epitope tagging for which useful mutants were available for genetic replacement. One of the best candidates was RPT4, which is encoded by two isoforms (RPT4a and RPT4b) with 96% amino-acid sequence identity. This level of identity is near average for shared sequence identity among RPT subunits (Book et al., 2010); RPT2a and RPT2b share the most identity at 99%, and RPT5a and RPT5b share the least identity at 93%. Based on the structure of the yeast 26S proteasome (PDB ID: 4CR2) (Beck et al., 2012), we expected that the N-terminus of RPT4 would be solvent accessible, thus making it an attractive candidate for attaching an exposed tag that would minimally disrupt particle assembly. Furthermore, we identified several T-DNA insertion mutations within RPT4a and RPT4b in the Col-0 background that might generate null alleles suitable for genetic replacement (Figure 1A). DNA sequence analysis of the *rpt4a/b* mutants deter-

mined the insertion sites to be at nucleotides 681, -278, and -24 for *rpt4a-1*, *rpt4a-3* and *rpt4a-4* and 151, 409, and 558 for *rpt4b-1*, *rpt4b-2*, and *rpt4b-3*, from the ATG start codon respectively. The *rpt4a-1* and *rpt4b-2* alleles were particularly notable because the T-DNA insertions sites disrupted the coding regions at 88 and 101 residues upstream of the AAA-ATPase cassette (Figure S1) for each polypeptide (Figure 1B). To assay whether these insertion alleles disrupted accumulation of the RPT4a or RPT4b transcripts, we performed RT-PCR analysis on RNA extracted from these seedlings. As shown in Figure 1B, seedlings homozygous for either *rpt4a-1* or *rpt4b-2* failed to accumulate the corresponding full-length transcripts (Primers P1 + P5, and Primers P6 + P9 respectively). However, a small level of amplification of the 5' region was observed with primers that were 5' to the insertion site in both *rpt4a-1* and *rpt4b-2* plants (Primers P1 + P2, and Primers P6 + P7) suggesting that 5' products may still be generated. Additionally, a small amount of product relative to wild type was also detected 3' to the insertion site for *rpt4a-1* suggesting some mRNA from this region can also accumulate. At the protein level, an immunoblot analysis of SDS-PAGE-separated proteins from *rpt4a-1*, *rpt4b-2* and wild type Col-0 seedlings showed reduction in RPT4 protein in *rpt4a-1* compared to wild type, but little to no change in *rpt4b-2* (Figure 1E). While RPT4a is the dominant isoform with 169 expressed sequence tags as compared to 39 for RPT4b ,

our data shows a more prominent decrease in RPT4a levels, as compared to RPT4b. There were no differences in RPT4 immunoreactivity between RPT4a and RPT4b (communications with Kwang-Hee Lee), therefore a more complex regulation of global RPT4a/b expression than anticipated may occur. Phenotypic analysis of the mutants showed that they were similar in overall morphology to wild-type when grown on agar plates for 10 days in a long day (LD) photoperiod (16 h light / 8 h dark) or when transferred to unfertilized soil and grown until 42 days old in a short-day photoperiod (8 h light / 16 h dark) (Figure 1C). Taken together these data suggest that either isoform RPT4 alone is not essential.

Plants containing the mutant alleles *rpt4a-1* and *rpt4b-2* were transformed with their respective 2XFLAG-tagged genomic variants and selected for stable transgene insertion. Expression was driven by the native promoter of each respective gene, and the constructs included the 5 UTR, codons for a 2XFLAG tag and the genomic region including introns that bracketed the RPT4a/b coding regions (see Experimental Procedures). The initial transformants were self-fertilized and the homozygous lines for both the *rpt4* T-DNA insertion and the transgene were identified in the Basta resistant progeny by genomic PCR. Immunoblot analysis of proteins extracted from RPT4a/b::FLAG-RPT4a/b-transformed plants detected accumulation of the FLAG-RPT4 proteins, which were larger than the corresponding native RPT4 polypeptides

when detected with anti-RTP4 antibodies, and could be detected with anti-FLAG antibodies (Figure 1E). The lower bands in the anti RTP4 immunoblot were likely RTP4b in the case of the *rpt4a-1* mutant, and RTP4a in the case of the *rpt4b-2* mutant. The transformed plants were phenotypically indistinguishable from wild type, implying that ectopic expression of either isoform bearing the 2X-FLAG tag was not detrimental to growth and development (Figure 1D). Taken together these data showed that individual mutants with defect in a single RTP4 isoform exhibit a phenotype that is similar to wild-type, and that transgenic plants expressing their respective FLAG-tagged RTP4 variants displayed no gross morphological defects.

Using the strategy first described by Book et al. (Book et al., 2010), we exploited the homozygous FLAG-RTP4a/b *rpt4a/b* seedlings to affinity purify 26S proteasomes via the RP. Briefly, liquid-grown seedlings were frozen at -80 oC and then ground in a mortar and pestle, extracted in a tissue extraction buffer (see Experimental Procedures), clarified by centrifugation, applied to the anti-FLAG affinity column, washed, and then gently eluted with FLAG peptide. To better stabilize CP/RP association, 20 mM ATP was included in the indicated buffers (Book et al., 2010; Liu et al., 2006). To enrich for just the RP (using FLAG-RTP4a/b) or the CP (using PAG1-FLAG) the same protocol was used except that ATP was omitted from the buffers. As a control, we also attempted to purify 26S proteasomes

from wild-type tissues. Under conditions where 20 mM ATP was included in the purification buffer (indicated with a + in Figure 2A), the FLAG-RPT4a/b affinity purified samples showed a comparable banding pattern after protein separation by SDS-PAGE as compared to the PAG1-FLAG affinity-purified samples suggesting a similar complement of proteins were present (Figure 2A). The only notable exception was the observation of a band that likely represented the FLAG-tagged PAG1 subunit (indicated with a black arrow) in the PAG1-FLAG samples (Book et al., 2010). The size-shifted RPT4 subunits seemed to be masked by other subunits and were not readily identifiable via silver staining alone. Importantly, the mock affinity-purified samples from wild-type tissue showed very few contaminating proteins as judged by silver stain, which was important for down-stream tandem mass-spectrometry (MS/MS) analyses. Overall, these affinity purification strategies based on either RPT4a or RPT4b were able to successfully purify 26S proteasomes from whole *Arabidopsis* seedlings.

When ATP was omitted from the purification buffer, there was clear enrichment for the CP in PAG1-FLAG samples, and clear enrichment for the RP in both the FLAG-RPT4a and FLAG-RPT4b samples, as judged by the size and intensity of the protein bands in Figure 2A. Immunoblot analysis of these samples showed clear enrichment for two CP subunits PAG1 and PBA1 in the CP-based samples affinity-

purified in the absence of ATP, as compared to the RP subunits RPT4, RPN5, and RPN12 (Figure 2B). There was also clear enrichment for the three RP subunits in the RP-based affinity purified samples when ATP was omitted from the purification as compared to the CP subunits that were immunoblotted. Taken together, these data show that 26S proteasomes can be easily purified based on a RP affinity tag, and that both isoforms for a single subunit will work. By omitting ATP from the buffer, we were able to enrich for subunits of the CP or the RP utilizing affinity purifications from PAG1-FLAG and FLAG-RPT4a/b expressing tissues, respectively.

3.5.2. MS/MS Analyses of RPT4a/b Affinity-Purified Proteins Identifies a Similar Complement of Proteins as PAG1-FLAG Affinity Purifications

Previous MS/MS analyses of affinity preparations based on PAG1-FLAG identified the fundamental subunits that make up the 26S proteasome, including the CP subunits (PAA-PAG), the CP subunits (PBA-PBG), the RP AAA-ATPases (RPT) (RPT1-6), and the RP non-ATPase subunits (RPN) (RPN1-3 and RPN5-12). Most subunits are encoded as a paralogous gene pair (named 1 and 2 for CP subunits, and a and b for RP subunits) and most isoforms were found to be incorporated in the complex (Book et al., 2010). We wanted to verify that our affinity purifications based on RPT4a and RPT4b also contained a similar set of proteins as in Book

et al (Book et al., 2010). To do this we used label-free quantitative tandem mass spectrometry (LFQ-MS/MS) to analyze samples obtained from affinity purifications for FLAG-RPT4a, FLAG-RPT4b, PAG1-FLAG, and a mock wild-type control, in the presence and absence of ATP resulting in MaxLFQ values (Cox et al., 2014) for each protein analyzed. Importantly, we identified all subunit isoforms previously identified by Book et al. (Book et al., 2010) showing that our purifications were comparable (Figure 3, CP Ring, CP Ring, RPT, and RPN plots). Additionally, we identified a suite of key accessory proteins found to be statistically enriched as compared to control at a false discovery rate (FDR) of 0.01 for both the CP and RP (Figure 3 Associated Proteins).

The CP and subunits show higher MaxLFQ values in the PAG1-FLAG samples as compared to either FLAG-RPT4a/b sample under both plus and minus ATP conditions (Figure 3, CP Ring, CP Ring). This suggests some free CP was purified even in the presence of ATP, and that the complexes purified were not exclusively RP-CP proteasomes. In the PAG1-FLAG samples, the CP and subunits showed higher MaxLFQ values when ATP was omitted from the affinity purification buffer as compared to when ATP was present (Figure 3, CP CP Ring, CP Ring). This increase is consistent with the preferential purification of the CP over the RP in the absence of ATP when using the PAG1-FLAG epitope tag.

Likewise, both affinity purifications that target the RP through FLAG-RPT4a/b, had higher MaxLFQ values for RP subunits as compared to the CP-based PAG1-FLAG samples under both plus and minus ATP conditions. When ATP was omitted from the purification, the FLAG-RPT4a/b samples had increased MaxLFQ values for RP subunits as compared to when ATP was included; however, this increase was considerably less than that observed for the CP subunits in the PAG1-FLAG samples (Figure 3, RPT, and RPN plots). FLAG-RPT4a samples had higher MaxLFQ values for RPT4a when compared to FLAG-RPT4b and PAG1-FLAG (Figure 3, RPT plot). Conversely, FLAG-RPT4b samples had higher MaxLFQ values for RPT4b when compared to FLAG-RPT4a and PAG1-FLAG (Figure 3, RPT plot). These data suggest that, as expected, RPT4a was enriched in the FLAG-RPT4a samples, and that RPT4b was enriched in the FLAG-RPT4b samples.

3.5.3. Statistical Analysis of Affinity Enrichments for FLAG-RPT4a Show Little to No Difference When Compared to FLAG-RPT4b

While mammals and yeast can form only a few types of proteasomes, Arabidopsis, with its many subunit paralogs, can potentially form a much wider array of proteasome isotypes. By comparing samples enriched for RPT4a and RPT4b using a Significant Analysis of Microarray (SAM) test at 0.01 FDR, we determined if any

additional subunit isoforms (or other associated proteins) are incorporated preferentially, which would suggest that *Arabidopsis* can form proteasome isotypes. Surprisingly, this statistical analysis revealed that only when ATP was omitted from the purifications were the samples statistically enriched for either RPT4a or RPT4b, respectively (Figure S2B). While RPT4 and RPT4b were enriched in the samples in which ATP was included in the purification, they were not statistically significant by the SAM test at 0.01 FDR (Figure S2A). Additionally, the 26S proteasome subunits have larger MaxLFQ values in the FLAG-RPT4a sample suggesting a slightly more efficient purification for the proteasome with the RPT4a isoform. Furthermore, no significant enrichment of any other subunit isoforms was observed. As such these data suggest that there are no major differences in protein composition between affinity purifications for FLAG-RPT4a or FLAG-RPT4b other than statistically enriching for each individual isoform when ATP was omitted from the purification. This indicates that plants may assemble their proteasome randomly with respect to isoform incorporation under the conditions tested.

3.5.4. Statistical Analyses of LFQ-MS/MS Data Show Affinity Preparations Targeting Either the CP or RP Specifically Enrich for Their Respective Sub-complexes

The LFQ-MS/MS data allowed us to perform a more formal statistical test for the enrichment of the RP over the CP in samples affinity purified from the RP subunits FLAG-RPT4a/b, and for the enrichment of the CP over the RP in samples affinity purified from CP subunit PAG1-FLAG. We estimated the distribution of RP and CP subunits in these samples using kernel density estimations (KDE) of MaxLFQ values as compared to our wild-type mock-affinity purifications as a negative control (Figure 4 below each volcano plot, and larger plots in Figure S3). As the KDE distributions were not normal, two non-parametric statistical tests, Kolmogorov-Smirnov (KS) and median test, were used to determine if there was specific enrichment of particular proteasome sub-complexes (RP vs CP) from the different affinity purifications. The KS test determines if the two distributions are different in shape or position, and the median test determines if the distributions medians are statistically different.

The KDE distribution for the CP as compared to the RP from samples affinity purified in the absence of ATP using PAG1-FLAG showed both the KS and the median p-values being much less than 0.01, at 1.74E-10 and 4.3E-10, respectively

(Figure 4 and S3 A). This suggests that the distributions are different, and that their medians were different with a greater than 20 fold increase in CP over RP. Conversely the KDE distributions for the samples affinity purified in the presence of ATP using PAG1-FLAG (Figure 4 and S3 B) overlap to a higher degree, with both the KS and Median p-values > 0.01 at 0.401 and 0.256, respectively. This indicates that in samples affinity purified in the presence of ATP that the RP and CP distributions were not significantly different, and that their means were similar. Although not statistically significant, a 1.5 median fold change was observed for the CP over the RP in samples affinity purified in the presence of ATP using PAG1-FLAG suggesting that there may be some CP enriched as compared to the RP. These statistical analyses showed that, as expected, we were able to successfully enrich for the CP over the RP when affinity purifying from PAG1-FLAG expressing tissue.

Conversely, in samples using the RP-based FLAG-RPT4a/b purified in the absence of ATP, the KDE distributions showed enrichment for the RP over the CP, with a fold change of 3.43 and 3.41 for the RPT4a and RPT4b subunits respectively (Figures 4 and Figures S3 C and E, respectively). Both the KS (3.303E-5 for RPT4a and 5.55E0-5 for RPT4b) and median p-values (7.11E05 for RPT4a and 7.11E-5 for RPT4b) were < 0.01 , suggesting that the two distributions were statistically different and had different medians. Again, contrasting this with the samples purified in the

presence of ATP using FLAG-RPT4a and FLAG-RPT4b (Figures 4 and Figures S3 D and F, respectively), the KDE distributions overlap with both the KS (0.017 for RPT4a and 0.058 for RPT4b) and median p-values (0.089 for RPT4a and 0.023 for RPT4b) were > 0.01 , suggesting that the RP and CP distributions are similar. Although not statistically significant, we still observed a two-fold-change enrichment in both FLAG-RPT4a/b samples purified in the presence of ATP for the RP over the CP. In summary, these KDE distributions and statistical analyses provide further support for enrichment of either the CP or the RP, as expected, when samples were affinity purified in the absence of ATP via PAG1-FLAG, or FLAG-RPT4a/b, respectively.

3.5.5. Statistical Analysis of Affinity Purifications Enriched for the RP or CP Identifies Proteasome-Associated Proteins and Putative Assembly Chaperones Specific to Each Sub-complex

The ability to affinity-enrich both the CP and the RP from *Arabidopsis* under gentle non-denaturing conditions without the need for high salt washes may enable identification of associated proteins, particularly those which may be more loosely bound. Additionally, targeting the RP or CP enables the identification of proteins that may be associating with either or both subcomplexes specifically. The previous data represented in Figure 3 showed the raw MaxLFQ values for some of

the relevant associated proteins; however, a formal statistical test against mock affinity purifications was important given the high number of proteins identified and quantified via LFQ-MS/MS. To better understand the suite of proteins that are statistically enriched in samples that target the CP or RP specifically we compared the LFQ-MS/MS data obtained from our affinity purifications against their respective wild-type controls using volcano plots and SAM tests at 0.01 FDR.

Affinity purifications enriching for the CP via PAG1-FLAG, performed either in the presence or absence of ATP (Figure 4A and 4B respectively) show statistical enrichment for two orthologs of the UMP1 CP assembly chaperone, UMP1a and UMP1b. The Arabidopsis genome encodes a third UMP1 ortholog, UMP1c; however, we have no experimental evidence that UMP1c associates with the proteasome, even though it has several tryptic peptides from an extended C-terminal domain that could differentiate it from UMP1a/b (Figure S4). UMP1a, and UMP1b are considerably higher expressed at 90 and 55 ESTs as compared to UMP1c which had only 7 ESTs , suggesting that UMP1c may be lowly expressed, or possibly a pseudogene.

Intriguingly, these samples also show statistical enrichment for a suite of putative CP assembly chaperones (Proteasome Biogenesis Associated Chaperone) PBAC1-4 (Figure 4A and 4B). Although their annotation in The Arabidopsis Infor-

mation Resource v10 (TAIR10) was incomplete, an InterProScan of their respective sequences supported these identifications (Table S3A). While PBAC1 was annotated as having a PBAC2 domain, the expectation value of 3.4E-6 was much lower than the expectation value of 1E-29 for the PBAC2 domains in the potential PBAC2 homologue. Additionally, PBAC1 and PBAC2 were previously identified by iterative PSI-BLAST with the human PAC1 and PAC2 orthologs in two separate studies (Kusmierczyk et al., 2011; Le Tallec et al., 2007) which aids in our assignment of PBAC1 and PBAC2. This PSI-BLAST based analysis was repeated using human PAC1-4, which identified PBAC1-4 as the top hits for these proteins in Arabidopsis (Table S3B). Protein sequence alignments for PBAC1-4 show that these proteins are conserved across plants but that they show little conservation with orthologs outside the plant kingdom with < 15% protein sequence identity (Figure S5-S8). All statistically significant interactors from PAG1-FLAG affinity purifications performed in both the presence or absence of ATP are available in Table S4.

As shown in Figure 4 C-F, UMP1a/b and PBAC1-4 were not significantly enriched in our affinity purifications targeting the RP, suggesting that these proteins interact with the CP specifically. However, affinity purifications that targeted the RP through either FLAG-RPT4a or FLAG-RPT4b showed statistically significant enrichment over controls for the putative orthologs of RP assembly chaperones

HSM3, NAS2, and NAS6 (Figure 4 C-F). Importantly, these putative RP assembly chaperones were also absent from our affinity purifications that targeted the CP, suggesting that these proteins are RP specific interactors. All statistically significant interactors from FLAG-RPT4a and FLAG-RPT4b affinity purified in the presence or absence of ATP are available in Table S5 and Table S6. Several accessory proteins are common to both CP and RP purifications including the alternate capping particle PA200 (Book et al., 2010), ECM29 that is involved in stabilizing the CP-RP association (Lehmann et al., 2010), and the recently identified proteasome inhibitor 31 related protein PTRE1, which is involved in modulating proteasome activity in response to auxin (Yang et al., 2016).

3.5.6. A Novel Plant Proteasome-Associated Protein, PAP1, Interacts with a Putative CP Assembly Chaperone

An intriguing plant-specific protein, named here Proteasome Associated Protein 1 or PAP1 was identified and significantly enriched in affinity preparations that targeted the CP (Figure 4). While PAP1 had no domains of known function, as determined by InterProScan (Table S3A), it did contain a C-terminal HbYX (hydrophobic amino acid, followed by a tyrosine, followed by any amino acid) motif. This HbYX motif is commonly found in proteins that intercalate between the rings of the CP

(Kusmierczyk et al., 2011). This motif is found in both the plant specific PAP1, and in the assembly chaperone PBAC1 (Figure 5A) as well as the conserved yeast assembly chaperones Pba1 and Pba2 (Kusmierczyk et al., 2011). However, Pba2/PBAC2 orthologs outside of yeast typically contain a C-terminal HbF motif, with the only notable exception being the moss *Physcomitrella patens* which instead contains a HbYX motif. To confirm whether PAP1 did indeed interact with the proteasome, we generated transgenic plants expressing HA-tagged PAP1 in the PAG1::PAG1-FLAG pag1-1 background. Figure 5B shows immunoblots of SDS-PAGE separated samples from an affinity purification of HA-PAP1 utilizing anti-HA beads. HA-PAP1 expressed at very low levels in transgenic plants (Figure 5B, input). However, upon immunopurification and detection with anti-HA antibodies we were able to detect HA-PAP1, the FLAG-tagged CP subunit PAG1, and the FLAG-less PAG1 product that likely resulted from cleavage of the FLAG tag during overnight incubation (Book et al., 2010). Low levels of both the CP subunit PBA1, and RPN1 subunit were also detected in these samples.

To identify which putative *Arabidopsis* assembly chaperones might interact, we performed combination a yeast two-hybrid (Y2H) interaction assay for all pairwise combinations of PBAC1-4, and PAP1 (Figure 5C). We hypothesized that PAP1 may be a CP assembly chaperone and that it may interact with one or more of the

other putative assembly chaperones. The Y2H showed pairwise interaction for PBAC1 and PBAC2, and pairwise interaction for PBAC3 and PBAC4 regardless of whether they were attached to the GAL4 activating or binding domains (Figure 5C). These interactions were consistent with how the CP assembly orthologs in other organisms interact (Murata et al., 2009). Taken together these data suggested these proteins might form heterodimeric pairs, like other CP assembly chaperone orthologs. Importantly, we saw no interaction from PBAC1 or PBAC2 to PBAC3 or PBAC4 pairs in this Y2H, suggesting these putative assembly chaperones form separate complexes that their orthologs in mammals and yeast. Interestingly, we observed PAP1 interacting with PBAC1, which suggested that it may be part of the plant assembly chaperone PBAC1/2 complex (Figure 5C).

Because these interactions were analyzed in yeast, which contain their own assembly chaperones that might interfere with interactions between plant subunits, we wanted to verify that PAP1 interacted with PBAC1 in planta. To do this we tested several pairwise interactions by bimolecular fluorescent complementation (BiFC) via transient expression in *Nicotiana benthamiana* leaves. Figure 5D shows confocal images of transiently expressed PAP1, PBAC1, PBAC2, and PBAC3 with either the N- or C-terminal half of YFP (NYFP and CYFP respectively) attached as a translational fusion to the N-terminus of each putative assembly chaperone.

As controls, PAP1 and PBAC1-3 were tested for auto-activation against NYFP or CYFP alone with no strong auto-activation being observed (Figure S10B). We again observed interactions between PBAC1 and PBAC2, regardless of their attachment to the N or C terminal halves of YFP based on their strong fluorescent intensity (Figure 5D). Additionally, we saw an interaction between PAP1 and PBAC1, showing that PAP1 can interact with PBAC1 in planta. Consistent with our Y2H data, there was no evidence of strong interaction between PAP1 and PBAC3 (Figure 5D).

3.5.7. Inhibition of the Proteasome Results in Formation of Distinct Proteasome Sub-Species Containing Putative CP Assembly Chaperones

Upon inhibition with the proteasomal inhibitor MG132, most proteasome genes are coordinately upregulated by a pair of NAC transcription factors, NAC53 and NAC78, as the plants attempt to synthesize more proteasomes in an effort to increase degradation capacity (Gladman et al., 2016). Additionally, after prolonged inhibition the complex becomes heavily ubiquitylated and is turned over via autophagy, likely as a clearing mechanism for inhibited complexes (Marshall et al., 2015). We reasoned that treatment with MG132 might lead to an increase in the population of newly assembling proteasomes, potentially allowing us to observe proteasome assembly intermediates. As such, proteasomes were purified in the

presence of ATP from PAG1::PAG1-FLAG pag1-1 tissue treated for 16 hours with or without MG132. A silver stain of SDS-PAGE-separated proteins affinity-purified from MG132-treated tissue showed successful purification of the 26S proteasome under these conditions (Figure 6A) with the only notable difference in the treated sample being an increased association of PA200, which was also observed by Book et al. (Book et al., 2010). An anti-ubiquitin immunoblot of the same samples (Figure 6B), showed extensive ubiquitylation of the particle consistent with data observed by Marshall et al (Marshall et al., 2015).

Non-denaturing native-PAGE of affinity-purified proteasomes were performed to determine if any novel complexes could be observed. Intriguingly, silver staining of samples affinity-purified from MG132-treated tissue revealed a smear below the canonical CP band (Figure 6B). Additionally, extra bands were observed above the canonical CP band, which may represent stalled intermediates with the alternate capping particle PA200 (Li et al., 2007). To identify the proteins contained in the bands shown in Figure 7B, we cut the corresponding gel slices, trypsin-digested the proteins, and then analyzed the resulting peptides via MS/MS. A LFQ-MS approach was used to quantify the proteins present in each sample (Gemperline et al., 2016). Proteins identified in gel slices obtained from both MG132-treated and untreated control samples were clustered based on relative abundance using

hierarchical clustering. The clusters were then visualized via a heat map showing that the RP and CP proteins cluster away from one another in these samples (Figure 6C), consistent with their separation by native-PAGE. Furthermore, an outgroup that includes the putative assembly chaperones PBAC1-4, Ump1, and PAP1, plus the accessory factors PA200, PTRE1, as well as ubiquitin, is found in the CP cluster (Figure 6C).

The proteins identified in gel slices below the canonical CP band in both MG132-treated and untreated control samples (Figure 7B) contained a large abundance of CP subunits, with very few subunits (Figure 6C). In addition to this set of subunits, the previously mentioned CP-specific putative assembly chaperones PBAC1-4 were also observed, as was the plant-specific PAP1 protein and UMP1b. PBAC3 and PBAC4 were only identified in the slice below the canonical CP band. Slices 1-5 contained mostly CP subunits with very few, if any, RP subunits (Figure 6C). Slice 2, above the canonical CP band, from both treated and untreated samples also contained PBAC1, PBAC2, and PAP1, suggesting that this band may represent an assembly intermediate of the plant CP. All CP subunits and the alternate capping particle PA200 were identified in slice 3, suggesting that this band represents CP capped with PA200. Slices 4 and 5 were observed only in proteasomes purified from MG132 treated tissue (Figure 6B, and Table S8), and interestingly, slice 4 contained

both the PA200 alternate capping particle, and the assembly chaperones PBAC1, PBAC2, and PAP1 along with most of the CP subunits, suggesting that this may be a PA200 containing assembly intermediate. Slice 5 likely represented a doubly capped PA200-CP-PA200 species, as no other putative assembly chaperones were present, and a strong signal for PA200, and the various CP subunits was observed.

In contrast to slices 1-5, most RP subunits were identified in slices 6-10. Based on size, slices 6-7 likely represented free RP complexes that have some associating subunits, while slices 8-9 likely represent doubly or singly capped RP-CP complexes (Book et al., 2010). It was difficult to discriminate between singly and doubly capped proteasomes in these native gel assays, and we did not observe discrete bands for these complexes. Slice 10 likely represented the holo-complex of the 26S proteasome (Book et al., 2010), and notably is the only slice in which the CP-RP stabilizing protein ECM29 was identified; however, it was detected at low abundance. Very low levels of the putative assembly chaperones PBAC1, PBAC2, and PAP1 were detected in these higher slices. Increased amounts of ubiquitin were detected in slices 1-5 for proteasomes that were affinity-purified from MG132-treated tissue, which also contained the putative assembly chaperones PBAC1, PBAC2, and PAP1. No ubiquitin receptors were identified in these slices, as expected, which suggests that ubiquitin is covalently attached; however we were unable to

identify any canonical ubiquitylation footprints (GlyGly) on peptides corresponding to any proteasome subunits in these samples. Taken together, these data suggest that the novel upper novel slices and lower smears observed might represent stalled assembly intermediates, and that they may be actively ubiquitylated and turned over to prevent the formation of aberrant complexes.

3.6. Discussion

Here we have shown that we can affinity purify proteasomes from plants expressing FLAG-tagged RP subunits RPT4a and RPT4b, and that these preparations are of sufficient purity to enable analysis by MS/MS. Using a LFQ-MS/MS strategy, we were able to show that these RP-based affinity purifications identify the core components of the 26S proteasome, and are comparable to our previously developed CP-based affinity purification. Interestingly the only major difference between the RPT4a and RPT4b purifications was the statistical enrichment for each respective isoform when purifications were performed in the absence of ATP. Even with this enrichment, no other isoforms or associated proteins were preferentially enriched. For RPT4a/b purifications performed in the presence of ATP, enrichment for each respective subunit was observed; however, this difference was not statistically significant. These data imply that plants do not assemble particular proteasome

subunit isoforms into specific sub-particles or isotypes at the young-seedling stage. While most plant proteasome subunits are encoded as paralogs, mammals instead have only a few subunit paralogs. 1, 2, and 5 are replaced in immune tissue with 1i, 2i, and 5i to form the immunoproteasome isotype (Nandi et al., 1996), and in thymus tissue the immunoproteasome contains 5t instead of 5i forming the thymus proteasome isotype (Murata et al., 2007). Outside of these catalytic subunits, the only other isoform switch involves replacing the 4 subunit with a variant only expressed in testes forming a testes specific proteasome isotype (Belote et al., 1998). While there is no evidence for *Arabidopsis* assembling proteasomes into isoform specific sub-types, duplication of the RPT family has also occurred in monocots, with rice having duplicated subunits for RPT1, 2, 4 and 5. This suggests that there may be some evolutionary benefit to having multiple copies of the RPT family (Shibahara et al., 2004). Interestingly, proteomic analysis of bran and callus rice tissue showed some small differences in isoform expression suggesting that analysis of different *Arabidopsis* developmental stages may provide additional insight into potential isoform-specific functions (Shibahara et al., 2004). In conclusion, we did not find proteasome isotypes at the young seedling stage that are specific to either RPT4 isoform. In other organisms, proteasome sub-types seem to be found in specific tissues or cell-types (Belote et al., 1998; Murata et al., 2007; Nandi et al.,

1996). Therefore, affinity purifications of proteasomes from different plant tissues and developmental stages may be an intriguing next step to explore the possible existence of plant proteasome isotypes.

In conjunction with our previously developed CP affinity purification (Book et al., 2010), this newly developed RP affinity purification allowed us to specifically enrich for either the CP or RP sub-complexes when purifications are performed in the absence of ATP. LFQ-MS/MS and statistical analyses of these affinity-purified proteasome sub-complexes enabled identification of sub-complex-specific PAPs. The most predominant RP-specific PAPs present in both RPT4a and RPT4b affinity-purified samples are likely orthologs of the RP assembly chaperones NAS2, NAS6, and HSM3. We were unable to identify any proteins that were likely orthologs for the RP assembly chaperone RPN14. In yeast, RPN14 and NAS6 bind together in a complex, and only double mutants show a strong growth defect in growth, suggesting that together they perform redundant functions (Funakoshi et al., 2009). These data suggest that that plant proteasome assembly has diverged from the yeast and mammalian pathways and may no longer require RPN14 to assemble the RP (Funakoshi et al., 2009). There are no obvious orthologs for RPN14 in the *Arabidopsis* genome (Book et al., 2010) which supports this hypothesis. The exact role of NAS2, NAS6, and HSM3 in plant proteasome assembly remains to

be determined; however, their specific association with the RP suggests that they may perform similar functions in RP assembly as in mammals and yeast. It will be intriguing to test if these RP assembly orthologs form similar modules with specific RP subunits like their yeast and mammalian counterparts (Nas2 with Rpt4 and 5, Nas6 with Rpt3, Hsm3 with Rpt1,2 and Rpn1, and Rpn14 with Rpt6) (Park et al., 2010). Several PAPs identified here were enriched in affinity purifications targeting both the RP and CP. Unsurprisingly, the alternate capping particle PA200 was enriched in our CP-specific purifications. This protein is known to interact with subunits of the CP (Ortega et al., 2005). However, we detect PA200 in our RP affinity purifications as well, which suggests that plants form a hybrid RP-CP-PA200 complex. This hybrid complex has been observed in yeast, and is speculated to aide in release of degraded substrate peptides (Ortega et al., 2005; Schmidt et al., 2005; Ustrell et al., 2002). In plants however, PA200 mutants display a remarkably wild-type phenotype (Book et al., 2010), which suggests that this hybrid complex is not essential in *Arabidopsis*. ECM29 was also found to be enriched as compared to mock controls in both CP and RP targeted purifications. ECM29 is known in yeast to bind the CP subunit PRE10 (named PAG1 in *Arabidopsis*), and the RP subunit Rpt5 . In yeast, ECM29 is one of the only PAPs known to bind both the RP and the CP (Lehmann et al., 2010), consistent with our observations. Ecm29

has been shown to inhibit the yeast 26S proteasome, and may be a checkpoint for damaged or aberrantly assembled complexes (Lehmann et al., 2010). While ECM29 likely performs a similar role in Arabidopsis, its function in plants remains to be investigated.

A recently characterized plant ortholog of proteasome inhibitor 31 (PI31), named Proteasome Regulator 1 or PTRE1 (Yang et al., 2016), was identified as a CP-specific PAP, providing confirmation of PTRE1's association with the CP of the 26S proteasome. However, the most predominant CP-specific PAPs identified here were putative orthologs of the CP assembly chaperones PBAC1-4, UMP1, and the plant specific protein PAP1. In our analyses, searches for PBAC1-4 in the Arabidopsis genome using assembly chaperones from yeast (*Pba1-4*) or humans (*PAC1-4*) failed with BLAST alone. This is not surprising given that orthologs from different kingdoms typically shared <25% identity, with fungi and animal forms of these chaperones typically sharing <15% sequence identity (Table S3-S6) despite having similar functions. However two independent groups identified orthologs to the assembly chaperones PBAC1 and PBAC2 using an iterative PSI-BLAST based approach with the human orthologs PAC1 and PAC2 (Kusmierczyk et al., 2011; Le Tallec et al., 2007). PSI-BLAST has several advantages over BLASTp in that it is typically more sensitive, and can be useful when identifying protein like PBAC1-4

that likely share structural homology but share little sequence homology (Altschul et al., 1997; Aravind and Koonin, 1999; Kusmierczyk et al., 2011; Yashiroda et al., 2008). We were able to identify PBAC1 and PBAC2 as the top *Arabidopsis* PSI-BLAST hits for human PAC1 and PAC2 (Table S3B), respectively, repeating the finding by the both Le Tallec et al. (2007), Kusmierczyk et al. (2011). Additionally, we were able to extend this analysis, with both PBAC3 and PBAC4 being top PSI-BLAST hits for human PAC3 and PAC4 (Table S3B). PSI-BLAST analysis with the yeast assembly chaperones Pba1-4 failed to identify the putative *Arabidopsis* orthologs PBAC1-4, suggesting that the plant CP assembly system may be different from the yeast system. Sequence-based InterProScan analyses showed that PBAC2-4 contained PAC2-4 domains (Table S3A), further supporting these assignments. InterProScan analyses showed that PBAC1 contained a PAC2 domain; however, the E-value for PBAC1 containing this domain is much lower than the assignment of this domain to PBAC2 (3.4E-6 vs 1.0E-29), and is much lower than the other assigned domains. Despite this disagreement, PBAC1 was the top hit in the *Arabidopsis* genome when searching via PSI-BLAST with human PAC1, as performed by ourselves and others, and for these reasons we have decided to use the name PBAC1 for this identified protein (Kusmierczyk et al., 2011; Le Tallec et al., 2007). Sequence alignments of PBAC1 and PBAC2 also showed conserved HbYX

and HbY/F motifs respectively that are also found in their human counterparts. Sequence analysis alone, however, is insufficient to conclude that these are assembly chaperones. There is strong divergence in sequence identity between the plant, yeast, and animal sequences (alignments and identity matrices in Figures S5-S8); therefore, additional characterizations will be necessary to determine what role if any they have in plant CP assembly. Despite this divergence, these bioinformatics analyses, along with prior literature (Kusmierczyk et al., 2011; Le Tallec et al., 2007), help support our identification of these CP-specific PAPs as PBAC1-4.

Phylogenetic analysis of the putative plant assembly chaperones orthologs PBAC1-4, and plant specific protein PAP1, showed that PBAC2-4 formed clades with their respective animal counterparts (Figures S11A and B); however, PBAC2 was the only protein to form a clear clade with all sequences from animals, yeast, and plants. PBAC1, 3, and 4 from yeast did not form clades with either the plant or animal sequences, suggesting that they are more different in sequence as compared to the plant and animal versions. In this analysis PBAC1 from yeast, mammals and plants was a strong exception in that it did not form a strong clade with any of their respective orthologs (Figures S11A). These data may not be surprising given the low sequence homology shared between yeast and animal PBAC1, with sequence identity at <13%. This low % identity makes phylogenetic analysis of

these putative PBAC1 orthologs particularly challenging, therefore we cannot draw strong conclusions about PBAC1 orthology based on this analysis alone. In yeast and mammals, Pba1/PAC1 and Pba2/PAC2 form a heterodimeric pair. Separately, Pba3/PAC3 and Pba4/PAC4 also form a heterodimeric pair. If plant PBAC1-4 are acting similarly, we would expect that PBAC1 should interact with PBAC2 and that PBAC3 should interact with PBAC4. Indeed, our Y2H analyses shows these interactions (Figure 5C) suggesting that they each form a heterodimer like their respective orthologs in yeast and mammals (Kunjappu and Hochstrasser, 2014). Additionally, in our Y2H we failed to detect interactions between these two putative heterodimeric pairs as both PBAC1 and PBAC2 failed to interact with PBAC3 and PBAC4, which suggests that these proteins form two separate complexes. These data were also re-capitulated in our BiFC analyses (Figure 5D) suggesting that these interactions occur in planta; however, the interaction between PBAC3 and PBAC4 was not tested via BiFC.

We were able to confirm the interaction of the CP-specific PAP1 protein with the proteasome by IP with an HA tagged version, identifying the CP subunit PAG1. The presence of a conserved HbYX motif in PAP1 (Figure 5A), which was also found in the putative assembly chaperone PBAC1, suggested that PAP1 might play a role in plant proteasome assembly. In yeast, the Pba1 and Pba2 proteins both have HbYX

motifs important for interacting with the CP ring; however orthologs of Pba2 in higher eukaryotes lack HbYX motifs, and instead have a HbY/F motif that lacks the last variable C-terminal amino acid (Kusmierczyk et al., 2011). Consistent with this, PBAC1 has a HbYX motif conserved in plants, and PBAC2 has a HbY/F motif conserved in all plants analyzed, except *Physcomitrella patens* (which has a HbFX motif, Figure 5A), suggesting the plant-CP-assembly pathway is more similar to that of higher eukaryotes than that of fungi. If PAP1 is functioning as a CP assembly chaperone it might be acting in concert with PBAC1, given their interaction (Figure 5C and D). In one scenario, PAP1 could replace PBAC2 in the PBAC1/2 heterodimer, or in an alternative scenario, PAP1 may be forming a novel trimeric complex with both PBAC1 and PBAC2. Despite our evidence that PAP1 contains a conserved HbYX motif, and interacts with the plant assembly chaperone PBAC1 via both Y2H and in planta by BiFC, we cannot rule out that PAP1 functions outside of plant proteasome assembly.

While an interaction network of these putative plant proteasome assembly chaperones provides some insight into how these proteins organize, our analysis of complexes affinity purified from PAG1-FLAG tissue treated with the potent and specific proteasome inhibitor MG132 shows the formation of novel sub-complexes that contain these putative chaperones. In general, two classes of species appeared

in the sample affinity purified from tissue treated with MG132 as analyzed by native-PAGE. The first is a smear running at a faster pace below the canonical CP species which was increased in intensity upon inhibitor treatment. The second class includes discrete species migrating more slowly than the canonical CP band (Figure 6B).

In the untreated sample, native-PAGE showed a very faint band (as opposed to a smear) below the canonical CP band (Figure 6B, labeled Half Barrel?). This may represent a proteasome half-barrel known as a 15S precursor complex that also typically runs below the CP band in native-PAGE analyses and contains partially assembled subunits along with UMP1 (Kock et al., 2015). Indeed, our MS/MS analyses of this faint band identified UMP1, subunits and the putative assembly chaperones PBAC1-4 and PAP1. While mostly subunits were present, some subunits were still identified at lower levels. In yeast, these 15S complexes lack the 7 subunit (Kock et al., 2015). We still detected the 7 subunit PBG1; however, it was present at a 50-fold lower level as compared to the CP band in our untreated samples (Table S5). Native-PAGE analysis of the affinity preparations from MG132-treated tissue showed a smear instead of the faint band seen in the untreated samples. MS/MS analyses of this smear showed that it also contained unassembled subunits, UMP1, and the suite of putative assembly chaperones PBAC1-4. We could

not observe a discrete band similar to the 15S species from yeast but this may be due to the fact that we are inhibiting the complex and are capturing a wider array of partially assembled products, or that these partially assembled products are less stable in plants as compared to yeast. Indeed, increased levels of ubiquitin were found in this fraction even though no ubiquitin binding proteins were detected (as expected), suggesting that these sub-species may be aberrant and are likely turned over, either via autophagy or other active proteasomes. PBAC3, PBAC4, and UMP1 are exclusively detected in the untreated and treated smears below the CP band, which is consistent with PBAC3 and PBAC4 preferentially binding free subunits at early stages of proteasome assembly, as established in both yeast and mammals (Kunjappu and Hochstrasser, 2014), and with UMP1 being degraded at later stages of proteasome assembly (Ramos et al., 1998).

The new species observed above the canonical CP band upon MG132 treatment contained large amounts of PA200, suggesting that these are likely singly- and doubly-capped versions of the CP (CP-PA200, and PA200-CP-PA200, respectively). The first band above the CP was previously identified as CP-PA200 by our lab (Book et al., 2010); however, this doubly capped complex and additional bands observed in affinity purified complexes from MG132 inhibited tissue have not been seen before in plants. Recently, other researchers have found that human cell

lines treated with the proteasome inhibitor bortezomib induced the formation of a species above the canonical CP band, which also contained PA200 (Welk et al., 2016) suggesting that inhibition of the complex increases CP-PA200 binding, and that this process is conserved between animals and plants. The function of increased PA200 binding to the CP is still unclear in both plants and animals. It could be that inhibition of the complex causes PA200 to associate and help release peptides from stalled proteasomes. Alternatively, given PA200's role in assembly (Savulescu and Glickman, 2011), it could represent increased biogenesis of the CP sub-complex in response to inhibition.

Intriguingly, high levels of PBAC1, PBAC2, and PAP1 were found in the native-PAGE bands directly above the canonical CP band (slice 2 Figure 6B and C) in samples affinity purified from both untreated and MG132 treated tissues. This suggests that PBAC1, PBAC2, and PAP1 may be acting at later stages of proteasome assembly, given their association with species that run higher than mature CP. While the assembly chaperones Pba1/PAC1 and Pba2/PAC2 in yeast and mammals typically do not associate with mature forms of the CP, the related archaeal assembly chaperones PbaA and PbaB are found loosely associated with mature CP when its active sites are bound with inhibitors (Kusmierczyk et al., 2011). The data gathered here suggest that PBAC1, PBAC2, and PAP1 can bind mature CP and suggest that

plant CP assembly at later stages may be more similar to archaeal proteasome assembly. Taken together these larger-sized bands may represent stalled assembly intermediates of the mature proteasome still carrying their assembly chaperones. Like the previously mentioned smears, increased levels of ubiquitin are found in these fractions even though no ubiquitin binding proteins were detected, suggesting that these sub-species may also be aberrant and are likely turned over. While the exact role of PAP1 in plant proteasome assembly, if any, still needs to be determined, the data presented here suggest that it may be involved in late-stage proteasome assembly, as it co-migrates with PBAC1 and PBAC2. Figure 7 illustrates a possible model of CP assembly that is consistent with the data collected here.

In summary, the data presented herein show that we were able to efficiently purify the 26S proteasome via both the CP and RP. The newly-developed RP-based purification allowed us to identify CP and RP specific PAPs, several of which may be involved in the assembly of their respective sub-complexes. The fact that the putative CP assembly chaperones were only identified in our CP-based affinity purifications, and that the putative RP assembly chaperones were only identified in our RP-based affinity purifications, suggest that only mature RP is purified with our CP affinity preparation, and that only mature CP is purified with our RP affinity purification. Our follow up analysis builds an interaction network of PBAC1-4

and PAP1, and identifies novel sub-species that contain these putative assembly chaperones when analyzing sub-complexes affinity purified from MG132-treated tissue. Taken together, these data suggest that UMP1, PBAC3, and PBAC4 may be performing a role in the early biogenesis of the CP, while PBAC1, PBAC2, and possibly PAP1 are playing a role in later stages of CP assembly. While we have been able to identify the likely players in the plant proteasome assembly pathways for both the RP and CP, additional functional analyses of these proteins will be needed to demonstrate their roles as putative assembly chaperones, and to fully elucidate the plant proteasome assembly pathway. Figure Legends Figure 1 *rpt4a-1* and *rpt4b-2* mutants rescued with flag tagged variants. (A) Organization of the Arabidopsis RPT4a and RPT4b genes with insertion points shown for T-DNA alleles of RPT4a and RPT4b. (B) RT-PCR analysis of the RPT4a and RPT4b transcripts in wild type and *rpt4a* and *rpt4b* mutant seedlings showing aberrant RPT4a transcripts in *rpt4a-1* and abnormal RPT4b mRNAs in *rpt4b-2* plants. (C) Homozygous *rpt4* mutants show a phenotype similar to wild-type control. The top panel shows plants grown for 10-days in a 16-hour photoperiod, while the lower panel shows plants grown for 42 days in an 8-hour photoperiod. (D) Replacing the wild-type RPT4a and RPT4b with their respective flag-tagged variants show plants with normal growth when grown for 7 days on GM plates and then grown in soil for two weeks.

(E) Immunoblots using anti-RPT4 and anti-FLAG antibodies confirm the presence of the tag and an increased size due to the FLAG tag, which is reactive to the anti-FLAG antibody. A loading control for histone-3 is shown.

Figure 2 Protein composition of affinity-purified proteasomes from FLAG-RPT4a-, FLAG-RPT4b-, and PAG1-FLAG-expressing plants and from mock wild-type controls in the presence (+) or absence (-) of 20mM ATP. (A) Silver stained gel of each preparation shows clear enrichment over the negative control (Col-0) with minimal contaminants, both in the presence and absence of ATP during purification. Omitting ATP from the preparations for the CP-based PAG1-FLAG pulldown shows a decreased amount of RP subunits, while both RP-based subunit pulldowns show a decrease in the amount of CP subunits. (B) Immunoblots for various proteasome subunits for both the RP and the CP show a clear enrichment of proteasome subunits as compared to the wild type mock affinity purification control. As shown in the silver stain, omitting ATP from the preparations in each affinity purification causes a decrease in association of the alternative sub-particle (decrease in RP for CP pulldowns, and decrease in CP for RP pulldowns).

Figure 3 Label-free quantitative MS/MS analysis of affinity-purified proteins in the presence (+) or absence (-) of ATP. (CP Alpha Ring) subunits are enriched in samples that were affinity purified via the CP subunit PAG1-FLAG. This enrichment

increases even further when ATP is omitted from the preparations as shown on the left side of the graph. (CP Beta Ring) subunits are enriched at a higher level when purified via the CP subunit PAG1-FLAG. This enrichment increases even further when ATP is omitted from the preparations as shown on the left side of the graph. (RPT Subunits) RPT subunits are enriched at a higher level when purified via RP subunits RPT4a or RPT4b. There are clear increases in the level of enrichment of the RPT4a subunit for the RPT4a affinity purifications, and clear increases in enrichment of the RPT4b subunit for the RPT4b affinity purifications. These respective increases are greater than the isoform unbiased PAG1 pulldowns. The respective enrichments of each isoform also increases when ATP is omitted, due to a decrease in the possibility of binding the alternative subunit isoform under ATP limiting conditions. (RPN Subunits) RPN subunits are enriched at a higher level when purified via RP subunits RPT4a and RPT4b. This enrichment increases even further when ATP is omitted from the preparations as shown on the left side of the graph. (Associated Proteins) Key proteins identified as proteasome interacting proteins are specifically enriched in pulldowns based on the CP subunit PAG1, and others are specifically enriched for pulldowns based on either RP subunits RPT4a or RPT4b. Please refer to Figure 4 for statistical analyses of CP and RP distributions in affinity-purified samples, and for statistical analyses that identify candidates as

associated proteins.

Figure 4 Volcano Plots comparing proteasome affinity purifications under both plus and minus ATP conditions relative to WT controls, allow identification of statistically significant interacting partners for the proteasome. Volcano plots comparing affinity purifications for PAG1-FLAG (A and B), FLAG-RPT4a (C and D), and FLAG-RPT4b (E and F) against their respective WT mock affinity purification control are shown. The top panels (A, C, and E) show comparisons for samples purified in the absence of ATP, while the lower panels (B, D, and F) show comparisons for samples purified in the presence of ATP. Two-dimensional Kernel Density Estimations (KDE) are shown for CP subunits in red and RP subunits in blue. Larger KDE plots are available in Figure S3. CP subunits are labeled as red squares, while RP subunits are labeled as blue squares. Proteins outside the dashed lines are statistically significantly differentially represented in purified samples relative to control as judged by a significant analysis of microarray or (SAM) test at 0.01 false discovery rate. Proteins likely involved in CP assembly are shown as green squares, while proteins likely involved in RP assembly are shown as yellow squares. Known proteasome associated proteins or PAPs are labeled as black squares, while unknown PAPs are labeled in grey squares. Proteins identified that did not reach SAM significance are labeled as grey circles.

Figure 5 PAP1 interacts with the proteasome and with putative assembly chaperone PBAC1. (B) C-terminal sections of protein sequence alignments for PAP1, PBAC1 and PBAC2, show conserved HbYX / HbY/F motifs underlined in red (full alignments in Figures S9, S5, and S6) (B) An anti-HA immunopurification of HA-PAP1 expressed in the PAG1-FLAG background shows affinity enrichment for both PAP1 and the FLAG-tagged PAG1 subunit of the CP. A very low level of PBA1 is also observed in the elution. (C) A yeast-two-hybrid analysis of putative plant assembly chaperones shows interaction between PBAC1 and PBAC2 independently of whether PBAC1 is used as bait or prey. Similarly, interaction is also detected between PBAC3 and PBAC4 independently of which protein is used as bait. Interaction between PAP1 PBAC1 is also observed. (D) Bi-molecular Fluorescent Complementation (BiFC) with indicated combinations of PAP1, PBAC1, and PBAC2 fused at their N-terminus with either N or C terminal halves of YFP (N-YFP, C-YFP). BiFC results confirm the results obtained via Y2H in (B) and show that these interactions also occur in planta between PBAC1 and PBAC2, and between PBAC1 and PAP1 in either orientation. A control is shown in this figure documenting a lack of auto activation for N-YFP and C-YFP. Additional controls combining N-YFP and C-YFP with their respective N and C terminal PBAC1, PBAC2, PBAC3, and PAP1 controls are shown in Figure S8 along with the full panel from (D) for comparison.

Figure 6 Protein composition of PAG1-FLAG affinity-purified proteasomes from tissues treated with and without MG132. (A) A silver stain of proteasomes affinity purified from Arabidopsis tissue treated for 24 hours with 50 µM MG132 (left panel). An increase in PA200 is observed in these samples as judged by silver stain. An anti-ubiquitin immunoblot shows increased ubiquitylation of the complex on the right panel. (B) Native-PAGE analysis of 26S proteasomes affinity purified from MG132-treated (+) and untreated control (-) Arabidopsis tissues (left panel). A different banding pattern is observed between MG132-treated and untreated controls, with a smear below, and two bands (4 and 5) above the canonical CP band. An extended gel is shown on the right to better visualize and cut out the indicated bands for mass spectrometry. Cut bands indicated in red were recovered. The proteins contained within these bands were trypsin-digested and subjected to MS/MS analyses. (C) A heat map showing LFQ MS/MS analysis (using dNSAF values calculated with Morpheus Spectral Counter (Gemperline et al., 2016)) of the gel slices shown in (B), which were hierarchically clustered using a Pearson's correlation, and ordered by gel slice. Clear grouping of the CP and RP is observed. Fractions containing the unassembled rings and putative assembly chaperones are seen from the native gel smears shown in (B). Fractions two and four contain the putative assembly chaperones PBAC1, PBAC2, and PAP1 as well as CP subunits

and likely represent pre-assembly intermediates of the CP. PBAC3 and PBAC4 are only seen in the smear.

Figure 7 Putative Model of 20S Proteasome Assembly in Arabidopsis. A model of 20S proteasome assembly from yeast and mammals adapted from (Murata et al., 2009).

Figure S1 Sequence Alignments of RPT4a and RPT4b. Sequence Alignments of Arabidopsis RPT4a RPT4b and related sequences was performed with ClustalW. Positions of T-DNA insertion mutants for both RPT4a and RPT4a were determined by sequencing T-DNA specific PCR products and are shown with black triangles. The AAA-ATPase domain is indicated with a black bar. The Walker A and B motifs contained in the AAA-ATPase domain are indicated in red. Species names are abbreviated as follows: At *Arabidopsis thaliana*, Pt *Populus trichocarpa*, Os *Oryza sativa*, Sb *Sorghum bicolor*, Zm- *Zea mays*, Pp *Physcomitrella patens*, Sm *Selaginella moellendorffii*, Dm *Drosophila melanogaster*, Sc- *Saccharomyces cerevisiae*.

Figure S2 Volcano Plots Comparing Protein Profiles of Proteasome Affinity Purifications Based on RPT4a or RPT4b subunits in the presence (+) or absence (-) of ATP. (A) In the presence of ATP, statistical analysis using significant analysis of microarray (SAM) test does not show enrichment for RPT4a in the RPT4a

pulldown and RPT4b in the RPT4b pulldown. (B) In the absence of ATP, statistical analysis using a SAM test does show enrichment for RPT4a in the RPT4a pulldown and RPT4b in the RPT4b pulldown, but shows no other associations with central proteasome subunits or associated proteins. This suggests the complex assembles in a random fashion with respect to isoform incorporation.

Figure S3 Kernel Density Estimates (KDE) of the CP and RP distributions from label-free quantitative mass spectrometry show enrichment for the CP when purified via PAG1-FLAG, and for the RP when purified via FLAG-RPT4a/b. The X-axis shows difference in CP (red) or RP (blue) protein levels between affinity-purified samples and their respective mock affinity purification control, while the Y-axis is unit less with the area of each density corresponding to a total area of 1. Fold change (FC) CP over RP, or RP over CP was calculated based on the median of these distributions. Two p-values, one for a median test, and another for a Kolmogorov-Smirnov test, are shown, which test for differences in median and overall distribution respectively. Cryo-EM structures of the sub-complexes likely to be enriched in each affinity purification are shown, with free CP (A) and RP (C and E) in the absence of ATP, and in the presence of ATP half capped and doubly capped proteasomes (B, D, and F) adapted from (Beck et al., 2012) (A) The CP is enriched over 20 fold when performing an affinity purification with ATP omitted

from the purification buffer as compared to the RP. (B) The CP is only enriched 1.5 fold when affinity purified in the presence of 20 mM ATP. (C) The RP as compared to the CP is enriched approximately 3.5 fold in samples that were purified via FLAG-RPT4a in the absence of ATP. (D) In the presence of 20 mM ATP this RP enrichment decreases to only 2 fold for RPT4a. (E) In the absence of ATP the RP is enriched approximately 3.5 fold in samples affinity purified from RPT4b. (F) Enrichment for the RP decreases to about 2 fold in the presence of 20 mM ATP.

Figure S4 Multiple Sequence Alignment for UMP1. A Clustal Omega (Sievers and Higgins, 2014) sequence alignment of UMP1 from plants, yeast, and animals visualized with the program BOXSHADE. A heat map showing each combination of protein percent identity with respective values shown in each cell obtained from the resultant clustalo identity matrix.

Figure S5 Multiple Sequence Alignment for PBAC1. A Clustal Omega (Sievers and Higgins, 2014) sequence alignment of PBAC1 from plants, yeast (PBA1), and animals (PAC1) visualized with the program BOXSHADE. A heat map showing each combination of protein percent identity with respective values shown in each cell obtained from the resultant clustalo identity matrix. Species names are abbreviated as follows: A.th *Arabidopsis thaliana*, Z.ma *Zea mays*, Os *Oryza sativa*, M.tr *Medicago truncatula*, P.pa *Physcomitrella patens*, S.ec- *Saccharomyces cerevisiae*

K.la Kluyveromyces lactis, E.go Eremothecium gossypii, D.ha Debaryomyces hansenii, D.me Drosophila melanogaster, X.la Xenopus Laevis, D.re Danio rerio, M.mu Mus musculus, H.sa, Homo sapiens.

Figure S6 Multiple Sequence Alignment for PBAC2. A Clustal Omega (Sievers and Higgins, 2014) sequence alignment of PBAC2 from plants, yeast (PBA2), and animals (PAC2) visualized with the program BOXSHADE. A heat map showing each combination of protein percent identity with respective values shown in each cell obtained from the resultant Clustal Omega identity matrix. Species abbreviations are listed in Figure S5

Figure S7 Multiple Sequence Alignment for PBAC3. A Clustal Omega (Sievers and Higgins, 2014) sequence alignment of PBAC3 from plants, yeast (PBA3), and animals (PAC3) visualized with the program BOXSHADE. A heat map showing each combination of protein percent identity with respective values shown in each cell obtained from the resultant Clustal Omega identity matrix. Species abbreviations are listed in Figure S5

Figure S8 Multiple Sequence Alignment for PBAC4. A Clustal Omega (Sievers and Higgins, 2014) sequence alignment of PBAC4 from plants, yeast (PBA4), and animals (PAC4) visualized with the program BOXSHADE. A heat map showing each combination of protein percent identity with respective values shown in each cell

obtained from the resultant Clustal Omega identity matrix. Species abbreviations are listed in Figure S5

Figure S9 Multiple Sequence Alignment for PAP1. A Clustal Omega (Sievers and Higgins, 2014) sequence alignment of PAP1 visualized with the program BOXSHADE. A heat map showing each combination of protein percent identity with respective values shown in each cell obtained from the resultant Clustal Omega identity matrix. Species abbreviations are listed in Figure S5

Figure S10 Full BiFC Panel and Controls. BiFC analysis showing both GFP channel and a merged bright-field and GFP channel. (A) A repeat of Figure 6D is shown for reference (B) Additional controls combining N-YFP and C-YFP with their respective N and C terminal YFP-fused PBAC1, PBAC2, PBAC3, and PAP1.

Figure S11 Phylogenetic Analysis of PAP1 and PBAC1-4. Bayesian phylogenetic analysis of PAP1 and PBAC1-4 from sequence alignments shown in Figure S5-S9. Trees were generated with MrBayes (Ronquist et al., 2012) and then visualized with FigTree. Nodes are color-coded with clade probabilities (probability of observing this grouping in the set of sampled trees), and a color key is shown on the right. Species names are abbreviated as follows: A.th *Arabidopsis thaliana*, Z.ma *Zea mays*, Os *Oryza sativa*, M.tr *Medicago truncatula*, P.pa *Physcomitrella patens*, S.ec-
Saccharomyces cerevisiae K.la *Kluyveromyces lactis*, E.go *Eremothecium gossypii*,

D.ha Debaryomyces hansenii, D.me Drosophila melanogaster, X.la Xenopus Laevis, D.re Danio rerio, M.mu Mus musculus, H.sa, Homo sapiens. (A) Phylogeny generated with all sequences from yeast, animals, and plants. Clear grouping is observed between the yeast, animal, and putative plant PBAC2 orthologs into a single clade. Grouping was observed for PBAC3 between the animal, and putative plant sequences; however yeast PBAC3 was not found in this clade. Clear grouping is observed for PBAC4 between the animal, and putative plant sequences; however yeast PBAC4 was not found in this clade. Animal, yeast, and plant PBAC1 did not clearly group, likely due to considerable sequence divergence. (B) Phylogeny generated when omitting the yeast sequences and only utilizing animal and plant sequences. Clear grouping is observed between the putative plant orthologs PBAC2-4 with their respective animal orthologs into separate clades. Animal PBAC1 does not group with the putative plant specific PBAC1 orthologs, likely due to considerable sequence divergence from animal PBAC1.

Table S1 Oligonucleotide Primer Sequences Used in this Study. A table of primers used throughout this chapter is provided for reference.

Table S2 Contaminants Removed from MaxQuants ProteinGroups.txt file. A list of common contaminants that were removed from MaxQuants ProteinGroups.txt file for the proteomics experiments involving PAG1-FLAG, and FLAG-RPT4a/b

affinity purified proteasomes. ATG Identifiers, Symbols and descriptions from The Arabidopsis Information Resource v 10 (TAIR10) are provided.

Table S3 Bioinformatic Analysis of PBAC1-4 and PAP1 Sequences. (A) Inter-ProScan (Jones et al., 2014) analysis of Arabidopsis protein sequences for PBAC1-4 with InterProScan Signatures from both PFAM and Panther databases with their respective expectation-values (E-values) are shown for reference. While PBAC1 (previously identified as a likely ortholog for PAC1 (Kusmierczyk et al., 2011)) contains a Proteasome Assembly Chaperone (PAC) 2 domain shown with a star, it is of considerably lower expectation-value (E-value) than PBAC2 which contains a Proteasome Assembly Chaperone (PAC) 2 domain. PBAC3 contains a PAC3 domain, and PBAC4 contains a PAC4 domain. PAP1 contains no domains of known function but does contain a conserved domain annotated in the Panther database (Mi et al., 2005). (B) Iterative PSI-BLAST Analysis was performed on NCBI's non-redundant (nr) database using the human PAC1-4 sequences using the default parameters. PBAC1-4 were the top hits in Arabidopsis for their respective PAC1-4 search sequences. ATG identifiers, RefSeq identifiers, Number of PSI-BLAST iterations, and expectation values (E-value) are listed for each analysis. PBAC1 required a second iteration to be identified, in which all sequences above the default expectation threshold were included in the second search.

Table S4 Proteasome associated proteins identified after PAG1-based affinity purification under both plus and minus ATP Conditions. A table of proteins identified as significantly enriched at FDR 0.01 in the indicated affinity preparations. Shown are the proteins ATG identifiers, descriptions from The Arabidopsis Information Resource version 10 (TAIR10), symbols from TAIR10, average MS/MS count, average sequence coverage, fold change compared to control using the MaxLFQ value from MaxQuant, and p-values obtained from the volcano plots in Figure 4A and B.

Table S5 Proteasome associated proteins identified after RPT4a-based affinity purification under both plus and minus ATP conditions. A table of proteins identified as significantly enriched at FDR 0.01 in the indicated affinity preparations. Shown are the proteins ATG identifiers, descriptions from The Arabidopsis Information Resource version 10 (TAIR10), symbols from TAIR10, average MS/MS count, average sequence coverage, fold change compared to control using the MaxLFQ value from MaxQuant, and p-values obtained from the volcano plots in Figure 4C and D.

Table S6 Proteasome associated proteins identified after RPT4b-based affinity purification under both plus and minus ATP conditions. A table of proteins identified as significantly enriched at FDR 0.01 in the indicated affinity preparations.

Shown are the proteins ATG identifiers, descriptions from the The Arabidopsis Information Resource version 10 (TAIR10), symbols from TAIR10, average MS/MS count, average sequence coverage, fold change compared to control using the MaxLFQ value from MaxQuant, and p-values obtained from the volcano plots in Figure 4E and F.

Table S7 dNSAF values determined by Morpheus Spectral Counter used to generate the heatmap in Figure 6C. Proteins as annotated in the heatmap are listed with each slice and treatment status as treated or untreated with MG132. dNSAF values as determined by Morpheus Spectral Counter (Gemperline et al., 2016) are listed for each slice. Table S8 Fold Change dNSAF values comparing Slice 1 and Smear from Figure 6. Fold change values were calculated for both treated and untreated samples by dividing the dNSAF values for Smear / Slice 1. #DIV/0! Is listed as a divide by zero error, in that those proteins were detected in the smear but not in Slice 1. Larger values were color coded by red in Microsoft Excel, and smaller values were color coded by blue.

3.7. Literature Cited

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–402.
- Aravind, L. and Koonin, E.V.** (1999). Gleaning non-trivial structural, functional

and evolutionary information about proteins by iterative database searches. *J Mol Biol* **287**: 1023–40.

- Beck, F., Unverdorben, P., Bohn, S., Schweitzer, A., Pfeifer, G., Sakata, E., Nickell, S., Plitzko, J.M., Villa, E., Baumeister, W., and Förster, F.** (2012). Near-atomic resolution structural model of the yeast 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 14870–14875.
- Belote, J.M., Miller, M., and Smyth, K.A.** (1998). Evolutionary conservation of a testes-specific proteasome subunit gene in *Drosophila*. *Gene* **215**: 93–100.
- Besche, H.C., Haas, W., Gygi, S.P., and Goldberg, A.L.** (2009). Isolation of mammalian 26S proteasomes and p97/VCP complexes using the ubiquitin-like domain from HHR23B reveals novel proteasome-associated proteins. *Biochemistry* **48**: 2538–49.
- Book, A.J., Gladman, N.P., Lee, S.S., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2010). Affinity purification of the *Arabidopsis* 26S proteasome reveals a diverse array of plant proteolytic complexes. *Journal of Biological Chemistry* **285**: 25554–25569.
- Book, A.J., Smalle, J., Lee, K.H., Yang, P., Walker, J.M., Casper, S., Holmes, J.H., Russo, L.A., Buzzinotti, Z.W., Jenik, P.D., and Vierstra, R.D.** (2009). The RPN5 subunit of the 26S proteasome is essential for gametogenesis, sporophyte development, and complex assembly in *Arabidopsis*. *Plant Cell* **21**: 460–78.
- Brukhin, V., Gheyselinck, J., Gagliardini, V., Genschik, P., and Grossniklaus, U.** (2005). The RPN1 subunit of the 26S proteasome in *Arabidopsis* is essential for embryogenesis. *Plant Cell* **17**: 2723–37.
- Cox, J., Hein, M.Y., Luber, C.A., Paron, I., Nagaraj, N., and Mann, M.** (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**: 2513–26.
- Cox, J. and Mann, M.** (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**: 1367–72.
- Dange, T., Smith, D.M., Noy, T., Rommel, P.C., Jurzitzza, L., Cordero, R.J., Legende, A., Finley, D., Goldberg, A.L., and Schmidt, M.** (2011). The Blm10 protein

- promotes proteasomal substrate turnover by an active gating mechanism. *Journal of Biological Chemistry* **286**: 42830–42839.
- Edelheit, O., Hanukoglu, A., and Hanukoglu, I.** (2009). Simple and efficient site-directed mutagenesis using two single-primer reactions in parallel to generate mutants for protein structure-function studies. *BMC Biotechnol* **9**: 61.
- Elsasser, S., Chandler-Militello, D., Muller, B., Hanna, J., and Finley, D.** (2004). Rad23 and Rpn10 serve as alternative ubiquitin receptors for the proteasome. *J Biol Chem* **279**: 26817–22.
- Elsasser, S., Gali, R.R., Schwickart, M., Larsen, C.N., Leggett, D.S., Muller, B., Feng, M.T., Tubing, F., Dittmar, G.A., and Finley, D.** (2002). Proteasome subunit Rpn1 binds ubiquitin-like protein domains. *Nat Cell Biol* **4**: 725–30.
- Farmer, L.M., Book, A.J., Lee, K.H., Lin, Y.L., Fu, H., and Vierstra, R.D.** (2010). The RAD23 family provides an essential connection between the 26S proteasome and ubiquitylated proteins in Arabidopsis. *Plant Cell* **22**: 124–42.
- Fatimababy, A.S., Lin, Y.L., Usharani, R., Radjacompare, R., Wang, H.T., Tsai, H.L., Lee, Y., and Fu, H.** (2010). Cross-species divergence of the major recognition pathways of ubiquitylated substrates for ubiquitin/26S proteasome-mediated proteolysis. *FEBS Journal* **277**: 796–816.
- Finley, D.** (2009). Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annu Rev Biochem* **78**: 477–513.
- Fu, H., Doelling, J.H., Arendt, C.S., Hochstrasser, M., and Vierstra, R.D.** (1998). Molecular organization of the 20S proteasome gene family from *Arabidopsis thaliana*. *Genetics* **149**: 677–92.
- Funakoshi, M., Tomko, R. J., J., Kobayashi, H., and Hochstrasser, M.** (2009). Multiple assembly chaperones govern biogenesis of the proteasome regulatory particle base. *Cell* **137**: 887–99.
- Gamborg, O.L., Miller, R.A., and Ojima, K.** (1968). Nutrient requirements of suspension cultures of soybean root cells. *Exp Cell Res* **50**: 151–8.
- Gelvin, S.B.** (2003). Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev* **67**: 16–37, table of contents.

- Gemperline, D.C., Scalf, M., Smith, L.M., and Vierstra, R.D.** (2016). Morpheus Spectral Counter: A computational tool for label-free quantitative mass spectrometry using the Morpheus search engine. *Proteomics* **16**: 920–4.
- Gladman, N.P., Marshall, R.S., Lee, K.H., and Vierstra, R.D.** (2016). The Proteasome Stress Regulon Is Controlled by a Pair of NAC Transcription Factors in Arabidopsis. *Plant Cell* **28**: 1279–96.
- Gonnet, G.H., Cohen, M.A., and Benner, S.A.** (1992). Exhaustive matching of the entire protein sequence database. *Science* **256**: 1443–5.
- Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., and Hunter, S.** (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**: 1236–40.
- Julio, S. and Jose, S.S.J.** (2006). *Arabidopsis Protocols*, Second Edition **323**.
- Kock, M., Nunes, M.M., Hemann, M., Kube, S., Dohmen, R.J., Herzog, F., Ramos, P.C., and Wendler, P.** (2015). Proteasome assembly from 15S precursors involves major conformational changes and recycling of the Pba1-Pba2 chaperone. *Nat Commun* **6**: 6123.
- Kunjappu, M.J. and Hochstrasser, M.** (2014). Assembly of the 20S proteasome. *Biochim Biophys Acta* **1843**: 2–12.
- Kusmierczyk, A.R., Kunjappu, M.J., Kim, R.Y., and Hochstrasser, M.** (2011). A conserved 20S proteasome assembly factor requires a C-terminal HbYX motif for proteasomal precursor binding. *Nat Struct Mol Biol* **18**: 622–9.
- Laemmli, U.K.** (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**: 680–685.
- Lander, G.C., Estrin, E., Matyskiela, M.E., Bashore, C., Nogales, E., and Martin, A.** (2012). Complete subunit architecture of the proteasome regulatory particle. *Nature* **482**: 186–191.
- Lasker, K., Förster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W.** (2012). Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 1380–1387.

- Le Tallec, B., Barrault, M.B., Courbeyrette, R., Guérois, R., Marsolier-Kergoat, M.C., and Peyroche, A.** (2007). 20S proteasome assembly is orchestrated by two distinct pairs of chaperones in yeast and mammals. *Molecular Cell* **27**: 660–674.
- Leggett, D.S., Glickman, M.H., and Finley, D.** (2005). Purification of proteasomes, proteasome subcomplexes, and proteasome-associated proteins from budding yeast. *Methods Mol Biol* **301**: 57–70.
- Leggett, D.S., Hanna, J., Borodovsky, A., Crosas, B., Schmidt, M., Baker, R.T., Walz, T., Ploegh, H.L., and Finley, D.** (2002). Multiple associated proteins regulate proteasome structure and function. *Molecular Cell* **10**: 495–507.
- Lehmann, A., Niewienda, A., Jechow, K., Janek, K., and Enenkel, C.** (2010). Ecm29 fulfills quality control functions in proteasome assembly. *Mol Cell* **38**: 879–88.
- Li, X., Kusmierczyk, A.R., Wong, P., Emili, A., and Hochstrasser, M.** (2007). beta-Subunit appendages promote 20S proteasome assembly by overcoming an Ump1-dependent checkpoint. *EMBO J* **26**: 2339–49.
- Lin, Y.L., Sung, S.C., Tsai, H.L., Yu, T.T., Radjacompare, R., Usharani, R., Fatimababy, A.S., Lin, H.Y., Wang, Y.Y., and Fu, H.** (2011). The defective proteasome, not substrate recognition function, is responsible for the null phenotypes of the Arabidopsis proteasome subunit RPN10. *Plant Cell* **23**: 2754–2773.
- Liu, C.W., Li, X., Thompson, D., Wooding, K., Chang, T.L., Tang, Z., Yu, H., Thomas, P.J., and DeMartino, G.N.** (2006). ATP binding and ATP hydrolysis play distinct roles in the function of 26S proteasome. *Mol Cell* **24**: 39–50.
- Livneh, I., Cohen-Kaplan, V., Cohen-Rosenzweig, C., Avni, N., and Ciechanover, A.** (2016). The life cycle of the 26S proteasome: from birth, through regulation and function, and onto its death. *Cell Res* **26**: 869–85.
- Marques, A.J., Glanemann, C., Ramos, P.C., and Dohmen, R.J.** (2007). The C-terminal extension of the beta7 subunit and activator complexes stabilize nascent 20 S proteasomes and promote their maturation. *J Biol Chem* **282**: 34869–76.
- Marshall, R.S., Li, F., Gemperline, D.C., Book, A.J., and Vierstra, R.D.** (2015). Autophagic Degradation of the 26S Proteasome Is Mediated by the Dual ATG8/Ubiqutin Receptor RPN10 in Arabidopsis. *Mol Cell* **58**: 1053–66.

- Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., Kitano, H., and Thomas, P.D.** (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res* **33**: D284–8.
- Murashige, T. and Skoog, F.** (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**: 473–497.
- Murata, S., Sasaki, K., Kishimoto, T., Niwa, S., Hayashi, H., Takahama, Y., and Tanaka, K.** (2007). Regulation of CD8+ T cell development by thymus-specific proteasomes. *Science* **316**: 1349–53.
- Murata, S., Yashiroda, H., and Tanaka, K.** (2009). Molecular mechanisms of proteasome assembly. *Nat Rev Mol Cell Biol* **10**: 104–15.
- Nandi, D., Jiang, H., and Monaco, J.J.** (1996). Identification of MECL-1 (LMP-10) as the third IFN-gamma-inducible proteasome subunit. *J Immunol* **156**: 2361–4.
- Ortega, J., Heymann, J.B., Kajava, A.V., Ustell, V., Rechsteiner, M., and Steven, A.C.** (2005). The axial channel of the 20S proteasome opens upon binding of the PA200 activator. *J Mol Biol* **346**: 1221–7.
- Paraskevopoulos, K., Kriegenburg, F., Tatham, M.H., Rösner, H.I., Medina, B., Larsen, I.B., Brandstrup, R., Hardwick, K.G., Hay, R.T., Kragelund, B.B., Hartmann-Petersen, R., and Gordon, C.** (2014). Dss1 is a 26S proteasome ubiquitin receptor. *Molecular Cell* **56**: 453–461.
- Park, S., Tian, G., Roelofs, J., and Finley, D.** (2010). Assembly manual for the proteasome regulatory particle: the first draft. *Biochem Soc Trans* **38**: 6–13.
- Pathare, G.R., Nagy, I., Bohn, S., Unverdorben, P., Hubert, A., Korner, R., Nickell, S., Lasker, K., Sali, A., Tamura, T., Nishioka, T., Forster, F., Baumeister, W., and Bracher, A.** (2012). The proteasomal subunit Rpn6 is a molecular clamp holding the core and regulatory subcomplexes together. *Proc Natl Acad Sci U S A* **109**: 149–54.
- Ramos, P.C., Hockendorff, J., Johnson, E.S., Varshavsky, A., and Dohmen, R.J.** (1998). Ump1p is required for proper maturation of the 20S proteasome and becomes its substrate upon completion of the assembly. *Cell* **92**: 489–99.
- Rappaport, J., Ishihama, Y., and Mann, M.** (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal Chem* **75**: 663–70.

- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P.** (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**: 539–42.
- Russell, J.D., Scalf, M., Book, A.J., Ladror, D.T., Vierstra, R.D., Smith, L.M., and Coon, J.J.** (2013). Characterization and quantification of intact 26S proteasome proteins by real-time measurement of intrinsic fluorescence prior to top-down mass spectrometry. *PLoS One* **8**: e58157.
- Sadre-Bazzaz, K., Whitby, F.G., Robinson, H., Formosa, T., and Hill, C.P.** (2010). Structure of a Blm10 complex reveals common mechanisms for proteasome binding and gate opening. *Molecular Cell* **37**: 728–735.
- Saveliev, S.V., Woodroffe, C.C., Sabat, G., Adams, C.M., Klaubert, D., Wood, K., and Urh, M.** (2013). Mass spectrometry compatible surfactant for optimized in-gel protein digestion. *Anal Chem* **85**: 907–14.
- Savulescu, A.F. and Glickman, M.H.** (2011). Proteasome activator 200: the heat is on. *Mol Cell Proteomics* **10**: R110 006890.
- Schmidt, M., Haas, W., Crosas, B., Santamaria, P.G., Gygi, S.P., Walz, T., and Finley, D.** (2005). The HEAT repeat protein Blm10 regulates the yeast proteasome by capping the core particle. *Nat Struct Mol Biol* **12**: 294–303.
- Schreiner, P., Chen, X., Husnjak, K., Randles, L., Zhang, N., Elsasser, S., Finley, D., Dikic, I., Walters, K.J., and Groll, M.** (2008). Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. *Nature* **453**: 548–52.
- Shibahara, T., Kawasaki, H., and Hirano, H.** (2004). Mass spectrometric analysis of expression of ATPase subunits encoded by duplicated genes in the 19S regulatory particle of rice 26S proteasome. *Arch Biochem Biophys* **421**: 34–41.
- Sievers, F. and Higgins, D.G.** (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* **1079**: 105–16.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., and Higgins, D.G.** (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**: 539.

- Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., and Cox, J.** (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* .
- Unverdorben, P., Beck, F., led, P., Schweitzer, A., Pfeifer, G., Plitzko, J.M., Baumeister, W., and Förster, F.** (2014). Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 5544–5549.
- Ustrell, V., Hoffman, L., Pratt, G., and Rechsteiner, M.** (2002). PA200, a nuclear proteasome activator involved in DNA repair. *EMBO J* **21**: 3516–25.
- VanderLinden, R.T., Hemmis, C.W., Schmitt, B., Ndoja, A., Whitby, F.G., Robinson, H., Cohen, R.E., Yao, T., and Hill, C.P.** (2015). Structural basis for the activation and inhibition of the UCH37 deubiquitylase. *Mol Cell* **57**: 901–11.
- Verma, R., Aravind, L., Oania, R., McDonald, W.H., Yates, J.R., Koonin, E.V., and Deshaies, R.J.** (2002). Role of the Rpn11 metalloprotease in deubiquitination and degradation by the 26S proteasome. *Science* **298**: 611–615.
- Vierstra, R.D.** (2009). The ubiquitin-26S proteasome system at the nexus of plant biology. *Nat Rev Mol Cell Biol* **10**: 385–97.
- Waskom, M.** (2016). Seaborn.
- Welk, V., Coux, O., Kleene, V., Abeza, C., Trumbach, D., Eickelberg, O., and Meiners, S.** (2016). Inhibition of Proteasome Activity Induces Formation of Alternative Proteasome Complexes. *J Biol Chem* **291**: 13147–59.
- Wenger, C.D. and Coon, J.J.** (2013). A proteomics search algorithm specifically designed for high-resolution tandem mass spectra. *J Proteome Res* **12**: 1377–86.
- Yang, B.J., Han, X.X., Yin, L.L., Xing, M.Q., Xu, Z.H., and Xue, H.W.** (2016). Arabidopsis PROTEASOME REGULATOR1 is required for auxin-mediated suppression of proteasome activity and regulates auxin signalling. *Nat Commun* **7**: 11388.
- Yang, P., Fu, H., Walker, J., Papa, C.M., Smalle, J., Ju, Y.M., and Vierstra, R.D.** (2004). Purification of the Arabidopsis 26 S proteasome: biochemical and molecular analyses revealed the presence of multiple isoforms. *J Biol Chem* **279**: 6401–13.

- Yao, T. and Cohen, R.E.** (2002). A cryptic protease couples deubiquitination and degradation by the proteasome. *Nature* **419**: 403–7.
- Yashiroda, H., Mizushima, T., Okamoto, K., Kameyama, T., Hayashi, H., Kishimoto, T., Niwa, S., Kasahara, M., Kurimoto, E., Sakata, E., Takagi, K., Suzuki, A., Hirano, Y., Murata, S., Kato, K., Yamane, T., and Tanaka, K.** (2008). Crystal structure of a chaperone complex that contributes to the assembly of yeast 20S proteasomes. *Nat Struct Mol Biol* **15**: 228–36.
- Yu, Y., Yu, Y., Smith, M., and Pieper, R.** (2014). A spinnable and automatable StageTip for high throughput peptide desalting and proteomics. *Protocol Exchange*.
- Zhang, X., Henriques, R., Lin, S.S., Niu, Q.W., and Chua, N.H.** (2006). Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat Protoc* **1**: 641–6.

Chapter A

AFFINITY ENRICHMENT OF THE POST-TRANSLATIONAL MODIFICATION UBIQUITIN

A.1. Summary

A.2. Introduction

A.3. Methods

A.4. Results

A.5. Discussion

A.6. Conclusions

A.7. Future Directions

COLOPHON

This document was typeset with L^AT_EX . It is based on the University of Wisconsin dissertation template created by William C. Benton (available at <https://github.com/willb/wi-thesis-template>).